Sensory Linguistics

Winter, Bodo

DOI:
10.1075/celcr.20

Sensory Linguistics:

Language, Perception, and Metaphor

Bodo Winter

Dept. of English Language and Applied Linguistics
University of Birmingham

Book manuscript submission to *John Benjamins*

Revisions / second submission

Für Dad

Acknowledgments

This book is an extension of the work I conducted for my PhD thesis at the University of California, Merced. Without the generous support of my advisor Teenie Matlock, none of this would have been possible. I am eternally thankful to her.

Much of the ideas presented here were developed during a visit to Madison, Wisconsin, where Marcus Perlman was a constant source of inspiration. I am forever thankful for the many great discussions we had together.

This book would not have been possible without the data collection efforts by Louise Connell and Dermot Lynott, who created the modality norms that form the focus of analysis. This book is essentially a love letter to their norms.

Tremendous thanks go to my father, who proofread every single chapter. I also want to thank Esther Roth at John Benjamins, as well as the series editors Kris Heylen and Ninke Stukker and the two anonymous reviewers. A special round of thanks also goes to Brittany Oakes, who has been an immense support by proof-reading the entire manuscript.

I want to thank Francesca Strik Lievers, Andre Coneglian, and Paula Pérez-Sobrino for providing input on an earlier chapter, and for shaping my thinking on "synesthetic metaphors." I have also learned a tremendous amount from Asifa Majid, who—together with Lila San Roque, Laura Speed, and Carolyn O'Meara—I want to thank for inviting me to the Perception Metaphor workshop at the Max Planck Institute for Psycholinguistics in Nijmegen. Much of the thinking in this book was further shaped by the questions and feedback I have received from attendees of my talks at RaAM Berlin, RaAM Odense, FTL3 Osijek, ICLC-14 Tartu and the International Corpus Linguistics Conference in Birmingham. I have also received great input on this topic by attendees of colloquium talks, and I want to thank Damian Blasi and the Zurich Center for Linguistics, Diane Pecher, René Zeelenberg, Beate Hampe, Márton Sóskuthy, Sarah Duffy, Martine Grice, Daniel Alcaraz Carrión, Patrycja Strycharczuk, and Grace Oh for giving me an opportunity to speak about this material. I also want to thank friends that have given input on the materials here—even if merely in casual conversations—including Christiane Schmitt, Timo Roettger, Roman Auriga, Julius Hassemer, and Thomas Kettig. Further thanks go to Bryan Kerster who supported me with Python and SQL in setting up COCA.

Finally, I want to thank my mum, Ellen Schepp-Winter, and the many other people who supported me during the time of writing this book, including Louis Hemming, Vincenzo de Salvo, George Lychnos, Jeannette Littlemore, Suganthi John, Nick Groom, Matteo Fuoli, Christian Mayer, Barry Dixon, Jim Digiambattista, Ruth Page, Adam Schembri, Amanda Patten, Tommy Seywert, and Pedro Miguel Martins. Special thanks go to the owners and staff of my favorite Birmingham places where significant parts of this book were written, including New Street Station's Joe and the Juice, Faculty Coffee, 200 Degree Coffee, Quarterhorse Coffee, and Tilt. I want to thank the Midland Outs Badminton group, my yoga teacher Anna Robottom, and the team of the Henrietta Street Gym for making life enjoyable during the time of writing (and no, Louis, I still don't know why animals can't talk).

Table of contents

Chapter 1. Sensory linguistics

1.1. Introduction

Humans live in a perceptual world. All of humanity's accomplishments, from agriculture to space travel, depend on us being able to interact with the world through seeing, feeling, hearing, tasting, and smelling. Everything we do, everything we feel, everything we know, is mediated through the senses. Because the senses are so important to us, it is not surprising that all languages have resources for talking about the content of sensory perception. Miller and Johnson-Laird (1976, p. 3) remind us that "word-percept associations are fundamental" to language. Rakova (2003, p. 34) says that we do not have words "just because it is nice for us to have them," but because they are "devices that connect us to the external world." In fact, without the ability to express perceptual content, language would be useless.

"Sensory linguistics" is the study of how language relates to the senses. It addresses such fundamental questions as: How are sensory perceptions packaged into words? Which perceptual qualities are easier to talk about than others? How do languages differ in how perception is encoded? And how do words relate to the underlying perceptual systems in the brain? The time is ripe for bringing these questions and many others together.

Research into the connection between language and perception has a long tradition in the language sciences. Among other things, researchers have looked at how many words there are for particular sensory modalities (e.g., Viberg, 1983), how frequently particular sensory perceptions are talked about (e.g., San Roque et al., 2015), and how metaphor (e.g., Ullmann, 1945; Williams, 1976) and iconicity (e.g., Dingemanse, 2012; Marks, 1978, Chapter 7) are used to achieve reference to the perceptual characteristics of the world. Yet, to my knowledge, no work has taken an integrated look at all of these topics together, along with other topics relating to perceptual language. That is what this book sets out to do.

The book is split into two parts. Part I focuses on *what* can be studied under the banner of "sensory linguistics." Part II focuses on the *how*. Whereas Part I is theoretical, focusing on the synthesis of existing research on language and the senses, Part II is empirical, providing a detailed analysis of English sensory adjectives and their linguistic and textual properties. To develop a comprehensive theory of the interaction between language and perceptual content, we need to focus on "those parts of the lexicon where dependence on perceptual phenomena is reasonably apparent" (Miller & Johnson-Laird, 1976, p. 119). In English, adjectives are the word class that is devoted to describing properties (see Givón, 2001; Murphy 2010), and what we may call "sensory adjectives" are those adjectives that are *about* sensory content (Diederich, 2015, p. 4; Lehrer, 1978). Because of this, sensory adjectives provide the optimal starting point for an investigation into sensory linguistics.

1.2. Contributions

There is already a lot of existing work on language and perception. Because of this, it is important to clarify what this book contributes. Together, there are three sets of contributions: descriptive, theoretical, and methodological contributions.

1.2.1. Descriptive contributions

On the descriptive side, this book characterizes the vocabulary of English sensory adjectives; how this vocabulary is composed and how it is used. With respect to describing the sensory vocabulary of languages, existing works have already looked at the language of color (e.g., Berlin & Kay, 1969), sound, and music (e.g., Barten, 1998; Pérez-Sobrino & Julich, 2014; Porcello, 2004), touch (e.g., Popova, 2005), temperature (e.g., Koptjevskaja-Tamm, 2015), pain (e.g., Lascaratou, 2007; Semino, 2010), taste and smell (e.g., Backhouse, 1994; Croijmans & Majid, 2016; Lee, 2015; Majid & Burenhult, 2014; Ronga, 2016). Few works have been published that compare two or more senses. For example, Dubois (2000, 2007) compares the language of sound with the language of smell, and Majid and Burenhult (2014) compare performance in a smell labeling task to performance in a color labeling task. Yet, there is little empirical work that encompasses all five common senses to attempt wide descriptive coverage (but see Miller & Johnson-Laird, 1976). Describing the sensory vocabulary of English is important for many reasons—not only for facilitating future theoretical work within linguistics and perceptual psychology, but also for applications in such domains as advertising, marketing, and food science (see, e.g., Diederich, 2015; Fenko, Otten, & Schifferstein, 2010).

1.2.2. Theoretical contributions

On the theoretical side, this book makes a number of contributions that are important for various aspects of linguistic research. I will begin by considering the notion of "ineffability," the difficulty of putting certain experiences into words (Levinson & Majid, 2014). Considering the limits of language poses deep questions about the nature of linguistic expressivity and whether it is truly the case that anything that can be thought can also be said (Searle, 1969).

Following this, I will deal with the topic of embodiment—what Wilson and Golonka (2013) call "the most exciting hypothesis in cognitive science right now" (p. 1). According to embodied approaches, we are "not just minds floating in the air" (Rakova, 2003, p. 18), and language is not just an abstract piece of software that can be instantiated by any physical system. Instead, embodied approaches see language and the mind as influenced by and deriving structure from bodily processes and sensory systems (e.g., Barsalou, 1999, 2008; Gallese & Lakoff, 2005; Gibbs, 2005; Glenberg, 1997; Lakoff & Johnson, 1999; Wilson, 2002). However, embodied approaches to cognition and language are not without their critics (Mahon & Caramazza, 2008). In a recent critique of embodiment, Goldinger and colleagues (Goldinger, Papesh, Barnhart, Hansen, & Hout, 2016, p. 964) rhetorically ask the question "What can you *do* with embodied cognition?" Their review suggests that many experimental findings within the cognitive sciences can seemingly do without invoking the notion of embodiment. In contrast to this, this book will show that it is not possible to do "sensory linguistics" without considering the notion of embodiment. As such, the empirical results presented in Part II provide an answer to Goldinger and colleagues' rhetoric question, showcasing one more thing that one can *do* with embodied cognition.

Considering embodiment also means that language scientists have to turn to other fields that study the senses in order to understand many of the patterns observable in language. As a result, sensory linguistics has to be an interdisciplinary endeavor: It has to look at the contributions of other fields that study the senses, such as psychology, neurophysiology, anthropology and philosophy. Any research that studies the connection between language and perception needs to consider both linguistic evidence and evidence from other fields. This book brings different strands of research from different disciplines together.

The final set of theoretical contributions relates to metaphor and polysemy, both of which are core topics in linguistics and cognitive science. It has been noticed by many scholars that sensory words such as *bright* and *rough* can be flexibly used to describe sensory experiences that are quite removed from what appears to be the core meaning of these words, such as in the expressions *bright sound* and *rough smell* (Ullmann, 1959; Strik Lievers, 2015; Williams, 1976). It is not clear how these so-called "synesthetic metaphors" fit into existing theoretical frameworks, such as conceptual metaphor theory (Gibbs, 1994; Kövecses, 2002; Lakoff & Johnson, 1980) and how they relate to psychological phenomena such as synesthesia (Ramachandran & Hubbard, 2001). Chapters 6 to 9 integrate the literature on synesthetic metaphors into the existing body of research on conceptual metaphors and the existing body of research on synesthesia. In doing so, I will also deconstruct the common view that regularities in how these metaphors are used are governed by a hierarchy of the senses.

The theoretical contributions presented in this book are not only relevant to the language sciences, but to the cognitive sciences more generally. After all, how language intersects with perception is one of the most fundamental questions we may ask about the language system and how it works together with other cognitive systems.

1.2.3. Methodological contributions

Finally, the book also makes a set of methodological contributions. Part II shows how sensory language can be studied objectively, using a mixture of human ratings and corpora. Many previous works on perceptual language have almost exclusively rested on the intuitions of linguists, with little quantification. Within the empirical study of sensory linguistics, some researchers have studied sensory language using experimental methods (e.g., Lynott & Connell, 2012; Speed & Majid, 2018) or field work (e.g., Floyd, San Roque, & Majid, 2018; San Roque et al., 2015). The study presented in Part II shows how far one can take an approach that uses ratings by naïve native speakers in conjunction with corpus data. Part II also demonstrates how a seemingly simple dataset of perceptual ratings for 423 sensory adjectives harbors a tremendous amount of theoretically relevant information—if subjected to the right statistical analyses. The methodological contributions of Part II are by no means limited to the study of sensory linguistics and have wide applicability in many areas of linguistics, especially in the quantitative study of semantics.

The book's empirical contributions also demonstrate how one can do sensory linguistics in a reproducible and open fashion. Reproducibility and open science are key topics in many different fields these days (Gentleman & Lang, 2007; Mesirov, 2010; Munafò et al., 2017; Peng, 2011), and the methodological approach outlined in Part II makes quantitative semantics and corpus linguistics more reproducible.

Some of the findings in Part II have already been believed to be true for a long time, but the evidence so far has relied too much on the intuitions of individual linguists. In part, the reproducible and objective approach advocated in this book will replicate what others have already done. In doing so, these ideas will rest on a firm quantitative footing, which facilitates future research in this domain. On the other hand, new methods naturally come together with theoretical adjustments. The methodological approach assumed in this book will lead to a number of incisive theoretical changes that affect foundational issues in sensory linguistics. For example, the new methods lead us to question the popular notion that there are five senses represented in language. It will also lead us to question the notion of a hierarchy of the senses.

1.3. Overview of the book

This book proceeds as follows. Chapter 2 discusses a potentially controversial choice that had to be made early on in the analysis presented here; namely, to adhere to what I will discuss as the "five senses folk model" of distinguishing the senses. Most work that has studied perceptual language up to this point has tacitly assumed a division of the sensory world into sight, touch, sound, taste, and smell. However, rather than taking this system as self-evident, this book will problematize this idea.

Chapter 3 introduces the many different semiotic strategies that can be used to express sensory content. Since later chapters in the book predominantly focus on analyzing sensory words, it is important to be clear from the outset that it is not only words that communicate sensory meaning. Instead, sensory language is characterized by semiotic diversity.

Chapter 4 introduces the notion of "ineffability," the difficulty of putting particular experiences into words. This notion was discussed at length in an important paper by Levinson and Majid (2014). This chapter draws from this work and extends it, making a number of distinctions that demand further attention.

Chapter 5 then moves to the topic of embodiment, reviewing the evidence for embodied semantics—the idea that words engage mental simulations of perceptual content. The evidence from embodied semantics will be used to argue for what I call the Embodied Lexicon Hypothesis, according to which the structure of the sensory world imprints itself into the structure of the lexicon and the way sensory words are used.

Chapter 6 is the first to deal with "synesthetic metaphors" such as *bright sound*, where an adjective primarily associated with sight (*bright*) is used to modify a noun primarily associated with sound. Such metaphors are often discussed in the context of the neuropsychological phenomenon of synesthesia. People with synesthesia experience vivid perceptual sensations in one sensory modality triggered by experiences in another concurrent sensory modality (such as seeing colors when hearing certain sounds). I will argue that any past invocations of synesthesia in the literature on "synesthetic metaphors" turn out to be red herrings: There is little connection between the two phenomena. This helps to clarify what "synesthetic metaphors" are in their own right.

Chapter 7 then argues that the term "synesthetic metaphors" is a misnomer because these "metaphors" are not all that metaphorical after all. Following Rakova (2003), I will present a literal (i.e., non-metaphorical) analysis of meaning extension in the domain of sensory words. This chapter also clarifies how crossmodal language is to be distinguished from other concepts in cognitive linguistics, such as metonymy and primary metaphor.

Chapter 8 then deals with asymmetries between the senses in meaning extension. For example, touch words are easily extendible to sound words (*rough sound, smooth melody*)—the reverse is not the case (*?squealing feeling, ?barking touch*). To account for such asymmetries, researchers have proposed a hierarchy of the senses. Chapter 9 deals more closely with what is supposed to ground this hierarchy, contrasting several different explanatory accounts. This completes Part I of the book.

Several of the theoretical ideas presented in Part I will recur in the empirical part of the book, Part II. The empirical part of the book is not meant to be an exhaustive study of all of sensory language, but rather a detailed case study demonstrating how one can go about *doing* sensory linguistics. Several important topics will be ignored, such as perception verbs and the history of sensory language, but the methodological approach used throughout these chapters can be used for any kind of investigation into the language of perception. Because the empirical approach adopted throughout the book is new and rests almost exclusively on the utility of a dataset of human ratings of sensory words, it is important to defend the methodological principles on which all analyses are based. This will be done in Chapter 10.

Chapter 11 focuses on how words can be classified according to sensory modalities. The issue of sensory classification is often glossed over in other research, even though many sensory words are highly multisensory, such as the word *harsh*, which can be used to express perceptual content from many different senses, such as *harsh sound*, *harsh taste*, *harsh smell*, and *harsh feeling*. This chapter on sensory classification also introduces the main dataset that will be used throughout the book, a set of modality ratings for 423 sensory adjectives that was collected by Lynott and Connell (2009).

Chapters 12 and 13 will present the first set of analyses of this dataset. Some of these analyses recapitulate what has been achieved by Lynott and Connell (2009) and Lynott and Connell (2013). These analyses mainly serve to familiarize the reader with the details of the set of sensory adjectives. However, several new analyses extend existing research by showing that sensory words occupy a "sweet spot" between sensory specialization and multisensoriality. I will furthermore present analyses which show that the dataset harbors particular clusters of words, such as words that relate to skin and temperature or words that relate to the spatial characteristics of the environment. In addition, I will show that correlations between the different perceptual ratings also suggest larger groupings, such as a part of the sensory lexicon that is devoted to expressing both taste content and smell content.

Whereas Chapters 12 and 13 look at the perceptual ratings in isolation, Chapter 14 looks at sensory words in context. Specifically, I will investigate how the senses are associated with each other in a corpus of naturally occurring language. This chapter shows that sensory words are used together in a way that reflects the structure of the senses in the real world; for example, taste and smell are closely aligned as perceptual modalities, and taste and smell words similarly stick together. Besides addressing several important topics in sensory linguistics, the methods used in this chapter provide a new quantitative look on core topics in corpus linguistics, such as the notion of "semantic preference."

Chapter 15 then relates the sensory words to other datasets. This chapter will highlight further linguistic asymmetries between the different senses; for instance, sight words are more frequent and more semantically complex, but sound and touch words are relatively more prone to harbor iconicity in their phonological structure. In a similar fashion, Chapter 16 shows that the senses differ in how they relate to the evaluative dimension of language. Taste and smell words are more evaluative than sight, touch, and sound words, which are relatively more neutral.

The final empirical chapter, Chapter 17, returns to the topic of synesthetic metaphors and brings the study of these expressions together with the topic of ineffability, presenting a new set of empirical analyses which question the notion of a monolithic hierarchy of the senses. In addition, I will show that the crossmodal use of sensory words is partially accounted for by linguistic factors discussed in previous chapters, such as the construct of "emotional valence" (see Chapters 5, 10, & 16) and the iconicity of sensory words (Chapter 15).

The final chapter of the book, Chapter 18, concludes by showcasing how the core themes discussed in the theoretical part of the book (Part I) are addressed by the various analyses presented in the empirical part (Part II). This chapter also discusses avenues for further research in sensory linguistics.

Together, the two parts of the book attempt to establish sensory linguistics as a field and showcase the utility of sensory linguistics in describing, and accounting for, important and widespread linguistic patterns that relate to the intersection of language and perception. The time is ripe to make "sensory linguistics" a subject in its own right, and this book is a first step toward this.

Part I.
Theory

Chapter 2. The five senses folk model

2.1. Introduction

Williams James (1891 [1890], p. 462) famously spoke of a "great blooming, buzzing confusion" of the senses. Humans are exposed to a complex "amalgam of sensory inputs" (Blake, Sobel, & James, 2004, p. 397). Yet, there is structure to perception. For example, color is perceived through the eye, which is a sensory organ that also simultaneously perceives spatial features such as distance and size, shape features, and motion through space. These perceptual features are correlated in our experience by virtue of being perceived through the same sensory organ and by presenting themselves to us at the same time. Humans can also perceive motion through sound, but sound experiences are principally dissociable from sight, such as when one's eyes are closed or when it is dark. Moreover, although sight may interact with the other senses in perception, it is associated with its own dedicated neural subsystem, such as the primary and secondary visual cortices. In this book, the term "sense" or "sensory modality" is used to refer to a subtype of perceptual experience that is associated with a dedicated sensory organ and its own cognitive machinery in the brain.

Traditionally at least, we recognize five senses: sight, hearing, touch, taste, and smell. These are sometimes called the "Aristotelian" senses (see, e.g., Sorabji, 1971). The analyses presented in this book are structured around the five senses of sight, hearing, touch, taste, and smell. A lot of research on language and the senses tacitly assumes the five senses model and takes it to be self-evident. However, it is not universally accepted that there are five senses, and it is not even clear that there are separate senses to begin with. Stoffregen and Bardy (2001, p. 197) say that "the assumption of separate senses may seem to be so self-evident as to be atheoretical," but instead, it carries "profound theoretical implications" (see also Cacciari, 2008, p. 431). This means that assuming five senses is an analytical choice that a linguist studying sensory language makes, and this choice may be problematic and, thus, has to be defended. This brief chapter will justify the analytical choice of adhering to the five senses folk model.

2.2. Issues with the five senses model

The idea that there are five senses, rather than fewer or more, is a culture-specific one. Classen (1993, p. 2) remarks that "even in the West itself, there has not always been agreement on the number of the senses," and anthropological research shows that not all cultures adhere to the five senses model (Howes, 1991). The cultural relativity of how the sensory world is partitioned is recognized explicitly by some researchers working on sensory language (e.g., Day, 1996), but it is often glossed over.

The five-fold way of carving up the sensory space furthermore does not correspond directly to everything we know from neurophysiology and perceptual psychology. Scientists recognize many subdivisions that do not fall neatly into the categories of sight, sound, touch, taste, and smell (Carlson, 2010, Chapter 7; Møller, 2012). For example, researchers recognize that pain is separate from other dimensions of touch: Pain perception is supported by underlying brain structures that are separate from regular touch perception (see, e.g., Craig, 2003; Tracey, 2005). Indeed, most researchers think of pain ("nociception") as a separate sense. Similarly, the so-called "vomeronasal organ" may be involved in constituting another sense that is different from the Aristotelian senses. This organ, partially separated from regular olfaction, is responsible for the perception of pheromones (see, e.g., Keverne, 1999). These are but two of many examples which fall through the cracks of the five senses model.

The existence of intense crossmodal interactions between the senses (see, e.g., Spence, 2011) poses further challenges for any attempt to classify the senses.

If there are separate senses, yet these senses interact both behaviorally and in terms of shared neural substrates, then it is not clear where to draw the boundary between two senses. For example, taste and smell are characterized by intense interaction (see Auvray & Spence, 2008; Spence, Smith, & Auvray, 2015). At what point do two senses interact so much as to call it one sense rather than two? Thus, multisensoriality poses another problem for categorizing the senses.

The question as to how many senses there are may be philosophically interesting, but it cannot be answered by any empirical data alone. It is not only a scientific question; it is also a definitional one. The problem is that there is no universally agreed set of criteria that could be used to differentiate the senses (Macpherson, 2011; Cacciari, 2008, pp. 430–431). We may define a sense with respect to what perceptual aspects of the world it specializes in, which was Aristotle's approach (Sorabji, 1971). We could also classify the senses with respect to what type of energy is involved, such as molecular energy (taste and smell), mechanical energy (touch, sound), or light energy (sight). Alternatively, we may define a sense as something that corresponds to a clearly recognizable body organ, such as the nose or the tongue, or we may define the senses with respect to the types of sensory receptors involved.

Not only are there many different criteria from which to choose, but each criterion itself is fuzzy. For example, what do we consider as a "body organ"? How are we to deal with distributed organs, such as the skin, or sensory systems that span the entire body, such as the internal senses? Do we treat neural tissue as being part of a sense? If so, the distinctions between the senses become even more messy, because the brain is massively interconnected. If we follow the receptor-based criterion, what divisions do we make? Should we treat mechanical perception and temperature perception as two separate senses because they are associated with their own receptors? But then, what about the many different types of mechanical receptors, with some receptors specializing in slow or fast vibrations, others in the perception of sustained touch, and still others in the perception of skin stretching? Shall we assign separate senses to each one of these receptors? These questions show the difficulty of establishing criteria for what constitutes a sense. Individuating the senses is a philosophically thorny issue that is at present unresolved (Casati, Dokic, & Le Corre, 2015; Macpherson, 2011) and perhaps even unresolvable. As McBurney (1986, p. 123) says, the senses "did not evolve to satisfy our desire for tidiness."

2.3. A useful fiction

Given this, is it not problematic that this book focuses on just five senses? The way the five senses folk model is used in this book is perhaps best seen as a "useful fiction." We may apply George P. Box's famous statement "All models are wrong but some are useful" here (Box, 1979, p. 2). The five senses model is a gross first-pass generalization that may form an adequate starting point for investigating language and perception. Miller and Johnson-Laird (1976, p. 16) say that the classification of words "by modality and [perceptual] attribute still represents the best psychological insight we have into the perceptual basis for their use." However, whereas Miller and Johnson-Laird (1976) assume the validity of the five-sense distinction, this book explores how far one can carry this model. To some extent, the proof is in the results, as the empirical chapters will show. Part II of this book will show that many important results can be derived from a five-fold division of the sensory world, even if this division is a deliberate abstraction. Dennett (2013, p. 31) says that oversimplification, such as is evidently the case when using the five senses model, allows scientists to "cut through the hideous complexity with a working model that is *almost* right, postponing the messy details until later."

One may also ask rhetorically: What other model could we possibly use? As stated above, there is no scientifically agreed list of the senses, nor do philosophers show any consensus. Given this, it is appropriate, for the time being, to use a culture-specific model. This book focuses on the English language, which is part of the cultural complex where people generally count five senses. When studying English, working with culturally endemic categories that are recognized by the speakers themselves is an advantage. Language is sometimes called a window into thought, but it can similarly be seen as a window into the folk model of sensory perception: Speakers "maintain conceptions about sensory perception as such; these conceptions are manifested in the linguistic expressions that designate perception" (Huumo, 2010, p. 50). There is much anthropological work on the fact that at least in the West and in modern times, the view that there are five senses is dominant (Classen, 1993; Howes, 1991). Given this, it makes sense to study how sensory words are used in English with respect to these cultural categories.

These cultural concerns are also connected to practical matters. First, it is desirable to achieve consistency with the large body of existing literature within linguistics that has already made interesting generalizations based on the five senses model. Second, the main dataset used in the empirical chapters (Part II) is a set of perceptual ratings collected from British native speakers by Lynott and Connell (2009). If we want to inquire native speakers' knowledge of sensory words in a reliable manner, it makes sense to stick to culturally endemic categories—asking naïve native speakers to make more fine-grained physiological distinctions may not be feasible.

2.4. Clarifications

It should be specified, however, what is regarded as a particular sense in this book and what is not. Following the folk model, the senses are each associated with one major sensory organ: the eye for vision, the ear for hearing (ignoring the vestibular system), the skin for touch, the tongue for taste, and the nose for smell (cf. Macpherson, 2011). Each one of these senses also has dedicated neural tissue. Although there is much crosstalk between different brain areas, and even though each brain area performs multiple tasks, one can identify visual cortex as primarily responsible for vision, the auditory cortex as primarily responsible for audition, the somatosensory cortex as primarily responsible for touch, the gustatory cortex as primarily responsible for taste, and the olfactory cortex (e.g., the piriform cortex) as primarily responsible for smell.

In this book, the word "touch" and the adjectives "tactile" and "haptic" are used in a deliberately broad fashion, as cover terms for everything that Carlson (2010, pp. 237–249) calls the "somatosenses," which describes those sensory systems that retrieve input from the skin, muscles, ligaments, and joints. This deliberately broad definition includes mechanical stimulation of the skin, thermal stimulation, pain, itching, kinesthesia, and proprioception. In focusing on the five senses folk model, these distinctions are deliberately ignored. The label "touch" is warranted because the bulk of the "touch words" dealt with in this book relates to the tactile exploration of surfaces, such as the words *rough*, *smooth*, *hard*, *soft*, *silky*, *sticky*, and *gooey*. A reason for including words such as *warm*, *hot*, *aching*, and *tingly* as belonging to touch is that this is how native speakers classify these words when having to them into the five senses folk model (Lynott & Connell, 2009). Thus, although there are more fine-grained distinctions when looked at from the perspective of physiology, we will follow the strategy of lumping the different somatosenses together for the time being. As some of the later chapters show (e.g., Chapter 13), more fine-grained distinctions can emerge out of an analysis that initially focuses on touch as a gross category.

The sensory modalities of taste and smell also warrant special attention: The folk model distinguishes these two senses, attributing the perception of flavor to the mouth and the tongue, even though flavor in fact arises from the interaction of taste and smell (as well as other sensory systems; see Auvray & Spence, 2008; Spence et al., 2015; see Chapters 13 and 14 for more details). However, when the terms "taste" and "smell" (and correspondingly "gustatory" and "olfactory") are used in this book, the folk sense is implied. Distinguishing taste and smell, at least initially, allows us to explore the relation between these two sensory modalities (see Chapters 13 and 14). Later analyses will explore to what extent taste and smell can be differentiated in language, or not.

Some researchers have made more nuanced distinctions between the senses in their linguistic studies (see Whitney, 1952; Williams, 1976; Ronga, 2016; Ronga, Bazzanella, Rossi, & Iannetti, 2012; Ullmann, 1945, 1959). Critically, different researchers have made different distinctions, making it difficult to compare across studies. Moreover, Healy (2017, p. 121) reminds us that "demands for more nuance actively inhibit the process of abstraction that good theory depends on." The five senses folk model is a considerable abstraction that, although not without its flaws, prevents us from falling into the nuance trap of making ever more fine-grained distinctions, many of which may not be reflected in language use. However, several of the chapters will show that concerns about oversimplifying the sensory world are not relevant because the analytical techniques used in this book can overcome the limitations of the folk model. Thus, although the choice of adhering to the five senses may at first seem theoretically limiting, this book demonstrates how these limitations can be broken down. The five senses model occupies a nice sweet spot, allowing us to make broad generalizations aimed at a high level of abstraction and generalization, while at the same time not being too constraining in the long run.

Chapter 3. Sensory semiotics

3.1. The sensory semiotic toolkit

To study how humans encode perceptual content, it "makes sense" to begin with semiotics, the study of meaning-making at its most general level. We can then explore how different semiotic strategies are used to talk about different aspects of our sensory world. Although this book focuses on sensory adjectives, we need to consider other semiotic strategies for communicating perceptual content. This is because sensory adjectives are part of a larger semiotic toolkit, and they are always chosen relative to this toolkit.

There is a "multiplicity of semiotic channels people use as they communicate in each other's presence" (Wilce, 2009, p. 31). Speakers use whatever semiotic means are available to them to achieve their communicative goals. The traditional way of structuring a discussion of semiosis is to adhere to the classic Peircean categories of icons, indices, and arbitrary symbols. Each of these sign types involves a different relation between the signal, such as a word or gesture, and an intended referent or meaning. The relations include perceptual resemblance (icons), direct connection (indices), and convention (arbitrary symbol; Clark, 1996, pp. 183–184).

Iconicity involves signaling meaning via resemblance, such as drawing. Indices are signs that establish meaning through a spatial, temporal, or causal relation (contiguity), such as smoke which indexes fire through causal contiguity (fire causes smoke) and spatial contiguity (smoke tends to be close to fire). Finally, symbols are best described in contrast to indices and iconicity: They establish meaning neither through resemblance nor through direct connection, but through arbitrary convention (de Saussure, 1959 [1916]). The only way to understand a symbol is to have learned its meaning. For example, there does not appear to be anything about the word form *purple* that gives away its meaning, and the German language actually has a completely different word for the same color, *lila*.

Clark (1996, Chapter 6) reconceptualizes Peirce's tripartite division of semiosis and speaks of "demonstrating" or "depicting", "identifying" and "describing-as." These terms serve to remind us that using icons, indices, and arbitrary signs are forms of communicative action. Iconic signs *depict*, indices *identify*, and arbitrary signs *describe-as* (e.g., the word *dog* describes a dog as a member of a conventionalized category in the English language).

Clark (1996) also emphasizes that the three forms of meaning-making "rarely occur in pure form" (p. 161). Iconicity, indexicality, and arbitrariness are not mutually exclusive, and they are frequently combined in what he calls "composite signals" (p. 161), which represent an "artful fusion of two or more methods of signaling" (p. 156). This idea is echoed by several other researchers who also view language from a wide lens: Enfield (2008) talks of "composite utterances" that combine gestures (often iconic or indexical) with speech. Kendon (2014, p. 3) speaks of a "semiotic diversity" in utterance production and highlights how utterances "always involve the mobilization of several different semiotic systems." Liddell (2003, p. 332) reminds us that "spoken and signed languages both make use of multiple types of semiotic elements."

I will now exemplify how sensory perceptions can be communicated using each of the three different semiotic strategies. The ability to encode perceptual content encompasses all three forms of semiosis. In addition to the three Peircean types of semiosis, it is useful to talk about "technical language" and "metaphor" as distinct categories as these are two means of communicating that have special relevance for encoding perceptual information. Technical language and metaphor are additional (derivative) means to express sensory content that are based on the three basic types of semiosis in some form or another, but they warrant discussion in their own right. The following provides a list of the different encoding strategies covered in this chapter (see also Barten, 1998; Portello, 2004):

1. iconicity (Chapter 3.2)
2. indexicality / source-based language (Chapter 3.3)
3. arbitrariness (Chapter 3.4)
4. technical language (Chapter 3.5)
5. metaphor (Chapter 3.6)

After discussing each encoding strategy in isolation, I will discuss common dimensions along which the different strategies vary (Chapter 3.7).

3.2. Depicting sensory perceptions with icons

Imagine a proud hunter who tells you about a massive hare she shot. In her description, she uses the adjective–noun pair *huuuuuuge hare* while simultaneously moving her hands apart, depicting the size of the animal. This multimodal expression is communicating perceptual content, in this case specifically relating to size. The manual component of this composite signal is an example of an iconic gesture, with gestural communication via iconicity being an important aspect of linguistic communication (Goldin-Meadow, 2003; Kendon, 2004; McNeill, 1992).

The phonetic lengthening of the adjective *huge* to *huuuuuuge* is another form of iconicity, what has been called either "iconic prosody" or "vocal gesture" (see Perlman & Cain, 2014; Perlman, Clark, & Johansson Falck, 2014). Such iconic prosody is perceptual imagery leaking through to vocal production. Perlman (2010) showed that English speakers reduce their speech rate when referring to distances that take longer to traverse (see also Perlman et al., 2014). Another set of studies has found that when speakers describe a visually presented moving dot, they increase their speech rate if the dot is moving faster, and they raise their voice pitch if the dot is moving upwards (Shintel & Nusbaum, 2007; Shintel, Nusbaum, & Okrent, 2006). All of these are examples of perceptual attributes (duration, distance, vertical position) being mapped directly onto particular phonetic parameters of speech production, such as duration and pitch.

These phonetic forms of iconicity are not encoded in the lexicon. “Iconic prosody” is dynamic and happens in the moment of speech production, where an already existing word is altered to convey a particular perceptual sensation. However, perceptual information can also form a stable part of a language’s vocabulary via phonological iconicity (Fischer, 1999; Schmidtke, Conrad, & Jacobs, 2014; see also Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015), or what others have called “sound symbolism” (Hinton, Nichols, & Ohala, 1994; see Ahlner & Zlatev, 2010 for a critique of the term “sound symbolism”). This type of iconicity is part of the phonological make-up of words; that is, specific phonemes or sequences of phonemes are directly linked to sensory meanings.

When it comes to the spoken word, the sensory domain of sound is a prime target for phonological iconicity (Dingemanse, 2012; Winter, Perlman, Perry, & Lupyan, 2017; see also Chapter 15), as exemplified by such onomatopoetic words as *bang* and *beep*. The term “onomatopoeia” refers to those cases of phonological iconicity that depict sound with sound. For example, words describing the sounds of music instruments may harbor a considerable degree of onomatopoeia (e.g., Patel & Iverson, 2003), and many bird names in the world’s languages are derived from bird sounds (Berlin & O’Neill, 1981). When it comes to sensory adjectives, some cases of onomatopoeia are difficult to spot:

Vickers (1984) discusses how the word *crispy* is used to describe food products that create high-pitched sounds when being chewed, compared to the word *crunchy*, which is used to describe food products that induce low-pitched chewing sounds. This contrast is mirrored by the word itself, with *crispy* having relatively more high frequency acoustic components than *crunchy* (see also Dacremont, 1995; Dijksterhuis, Luyten, de Wijk, & Mojet, 2007).

The following excerpt from a written review of the video game *Thumper* provides a particularly lively modern demonstration of onomatopoeia. In this review (from kotaku.com, http://kotaku.com/thumper-is-the-best-kind-of-music-game-1787670750, accessed August 18, 2017), the video game journalist Kirk Hamilton describes the game's unique interaction between sound and gameplay as follows:

> You'll constantly be alternating between listening and executing. Glowing gems become the kick drum; hard turns become the handclap snare. *BOOM… CLACK! … BOOM-CLACK!* You burst through barriers and they play a hi-hat fill. *Dat-da-dat-da-dat BOOM... CLACK!* (emphasis in original)

In the next paragraph, he provides the following description:

> The beat is spartan and precise, marching forward at a robotic medium tempo. Harmonies join the soundscape as frantic undulations rather than coherent chord progressions. It all channels the downbeat-heavy chug of a freight train: *CHUG-a-chug-a CHUG-a-chug-a CHUG-a-chug-a CHUG-a-chug-a*.

The italicized expressions exemplify onomatopoeia. Interestingly, these onomatopoetic forms are not integrated into sentences. The writer furthermore used italics to highlight the fact that these expressions are separate from the rest of the review. This fits the empirical observation that iconicity, when it is used expressively to convey vivid perceptual content, often lacks grammatical integration (Dingemanse & Akita, 2017).

But what if the sensory target domain is not sound? Phonological iconicity can also depict sensory meanings that are not directly related to sound. For example, in size sound symbolism, high and front vowels, such as /i/, are associated with small objects or animals; low and back vowels are associated with large objects or animals (Berlin, 2006; Diffloth, 1994; Fitch, 1994, Appendix 1; Marchand, 1959, p. 146; Ohala, 1984, 1994; Sapir, 1929; Thompson & Estes, 2011; Tsur, 2006, 2012, Chapter 11; Ultan, 1978). Phonological iconicity has also been established for other sensory qualities, including visually perceived speed of motion (Cuskley, 2013), luminance (Hirata, Ukita, & Kita, 2011), taste (Crisinel, Jones, & Spence, 2012; Gallace, Boschin, & Spence, 2011; Ngo, Misra, & Spence, 2011; Sakamoto & Watanabe, 2016; Simner, Cuskley, & Kirby, 2010), and texture (Etzi, Spence, Zampini, & Gallace, 2016; Fontana, 2013; Fryer, Freeman, & Pring, 2014; Sakamoto & Watanabe, 2018).

Probably one of the most explored sensory target domains with respect to phonological iconicity is the domain of shape. In a highly seminal paradigm spanning decades of research, researchers have explored how made-up words such as *kiki* and *takete* are associated with spikier and more angular shapes than *bouba* and *maluma* (Ahlner & Zlatev, 2010; Bremner, Caparos, de Fockert, Linnell, & Spence, 2013; Davis, 1961; Fischer, 1922; Köhler, 1929; Kovic, Plunkett, & Westermann, 2010; Maurer, Pathman, & Mondloch, 2006; Monaghan, Mattock, & Walker, 2012; Nielsen & Rendall, 2011, 2012, 2013; Ramachandran & Hubbard, 2001; Usnadze, 1924; also see Cuskley & Kirby, 2013, pp. 885–888). The output of this long line of research suggests that speakers of many languages, including young children, reliably associate the speech sounds of made-up words with shape characteristics (but see Styles & Gawne, 2017). However, research on the *kiki/bouba* phenomenon has so far focused on nonce words without establishing whether angularity as a perceptual property is actually encoded in the perceptual lexicons of natural languages.[1]

Ideophones, otherwise known as expressives or mimetics, are a special class of “marked words that depict sensory imagery” (Dingemanse, 2012, p. 654). Dingemanse (2018) says that ideophones are words that “show rather than tell” (p. 10) and whose meanings belong to “the domain of sensory imagery, evoking all sorts of perceptions and inner sensations” (Dingemanse, 2013, p. 143). Elsewhere, Dingemanse (2018) said that “a key feature of ideophones is that they evoke sensory qualities like motion, manner, texture, and colour” (p. 15). Ideophones can thus be seen as a specialized vocabulary for the depictive expression of perceptual content.

Some languages have thousands of ideophones, such as Japanese, which has ideophones such as *sara-sara* for smooth surfaces, *zara-zara* for rough surfaces, *puru-puru* for soft surfaces, *kachi-kachi* for hard surfaces, *gorogorogoro* for ‘rolling’, and *pikapika* for ‘shiny’ (Kita, 1997; Sakamoto & Watanabe, 2018; Watanabe, Utsunomiya, Tsukurimichi, & Sakamoto, 2012, p. 2518). Nakagawa (2012) describes a group of 32 “food texture verbs” in the Khoe language G|ui that describe mouthfeel by mimicking chewing sounds. These verbs include ideophones such as *χárù χárù* ‘fresh tsamma melon’ and *ts*^{h}*áǹ ts*^{h}*áǹ* ‘tender fillet meat’.

Dingemanse (2013, p. 148) recalls a situation in which a speaker of the African language Siwu (spoken in Ghana), describes what will happen if he sets fire to two small portions of gunpowder. The speaker utters the equivalent of *It'll go shû shû*, where *shû* is an ideophone that depicts a flaring event via a "sizzling sound" and short duration. While the speaker does this, he moves both hands quickly upwards, depicting the flaring event simultaneously via gestural iconicity. This is an example of a composite multimodal signal, and corpus studies show that ideophones often co-occur together with gestures (Dingemanse & Akita, 2017).

There are two constraints to the expression of sensory content via iconicity. First, iconic expression is always "selective" (see Clark & Gerrig, 1990; Hassemer & Winter, 2018); that is, only parts of the perceptual whole are represented, always at the expense of other parts. The resemblance between form and intended meaning is never perfect. Second, iconic expression is limited by the affordances of a language. Iconicity has to make do with what sounds and sound patterns a language makes available (see Styles & Gawne, 2017). Together, these two constraints explain why there are differences between languages even for onomatopoeia, the most direct form of iconic expression (Ahlner & Zlatev, 2010, p. 312; Marchand, 1959, pp. 152–153). The sound of a rooster, for example, is *cock-a-doodle-doo* in English and *kikeriki* in German. This is because different languages make different choices about what to iconically represent, and different languages make use of different sound inventories. The same applies to the examples seen in the video game review above. An expression such as *BOOM-CLACK!* is highly onomatopoetic, but it only represents the intended sound concept partially, and only via using a constrained set of English sounds.

Regardless of these limitations, it should be emphasized that iconicity, more so than any other semiotic strategy, is directly connected to the sensory world. Fittingly, Dingemanse (2011, p. 299) calls ideophones "the next best thing to having been there." In fact, Marks (1978, Chapter 7) likens phonological iconicity to a form of synesthesia, a union between the auditory-acoustic sense (speech sounds) and sensory meaning (see also Ramachandran & Hubbard, 2001; Sidhu & Pexman, 2017). Spence (2012) highlights the similarity between iconicity and crossmodal perception. Lockwood and Dingemanse (2015) review the many experimental effects that link particular perceptual sensations to speech sounds. Winter et al. (2017) showed that iconicity is more likely to occur on word forms if the words encode perceptual as opposed to abstract content (see also Sidhu & Pexman, 2018). Perlman, Little, Thompson, and Thompson (2018) replicated and extended this finding, showing that perceptual words are more iconic not only in English, but also in Spanish, American Sign Language, and British Sign Language. In fact, we may say that genuine iconicity has to be sensory in order to warrant the label "iconicity" because without perceptual content, there is nothing of which to be iconic. Abstract concepts devoid of sensory content are inimical to iconicity (Lupyan & Winter, 2018).

It has to be emphasized, however, that some perceptual qualities are more prone to iconic expression than others. Dingemanse (2012) reviews crosslinguistic evidence suggesting that ideophones most frequently encode sound, followed by movement, visual patterns, other sensory perceptions, and, finally, inner feelings and cognitive states. Winter et al. (2017) analyzed 3,000 English words rated for iconicity, finding that sound and touch words are the most iconic, followed by taste and smell words. In their data, sight words were found to be the least iconic. Perlman et al. (2018) furthermore showed that relatively more tactile words were relatively more iconic in both spoken languages (English and Spanish) and signed languages (American Sign Language and British Sign Language). Words strongly associated with sound were found to be more iconic, but only in spoken languages. For olfaction, gustation, and vision, Perlman et al. (2018) obtained negative correlations between perceptual strength and iconicity. This clearly shows that depiction as a semiotic strategy is more available for some sensory content than others, and this availability depends on the modality of expression (speaking versus signing). However, some perceptual qualities, such as color, are difficult to express iconically in any modality (Perlman et al., 2018).

3.3. Identifying perceptual qualities with indices

One of the most prototypical examples of an indexical sign is a pointing gesture, where there is a direct (projected) spatial relation to the signified object or person (Clark, 2003; Kendon, 2004, Chapter 11). Clearly, points can be used to identify perceptual aspects of the world around us. Points are, however, generally not about perceptual qualities, but rather about objects in the environment. In some constrained contexts, a point may actually be about a perceptual impression. For example, we may consider a situation with two bowls of water, one hot and one cold. In this case, a point could be used to signal temperature indirectly by pointing to either one of the bowls (example inspired by Dingemanse, 2013, p. 148). However, this point is only about temperature if the communicative context is right—the point's primary target is the object, not the perceptual impression.

Humans can identify a perceptual quality indirectly via referring to its source, even in its physical absence. This is what is sometimes called a "source-based strategy" or "source-based language" (e.g., Croijmans & Majid, 2016; Majid, Burenhult, Stensmyr, de Valk, & Hansson, 2018), exemplified by such expressions as *It sounds like a blender*, *It tastes like chicken*, or *It smells like kimchi*. At first sight, such expressions appear to be quite different from pointing gestures; however, there are two important similarities: First, both pointing gestures and source-based descriptions identify perceptual qualities only indirectly, via an object. Second, both communicative actions strongly rely on the speaker's "common ground" (Clark, 1996); in one case, the visual common ground; in another, shared background knowledge.

The similarity between pointing and source-based language becomes apparent when looking at the limitations of source-based language. Because the perceptual quality is not expressed directly, the intended perceptual quality has to be inferred when physical or verbal points are used (compare Levinson & Majid, 2014, p. 411; see also Fainsilber & Ortony, 1987, p. 241; Holz, 2007, p. 187). For example, the description *It tastes like kimchi* could refer to the dish's pungent and spicy properties, the carbonated taste, or the fishy or garlicky overtones that some forms of kimchi have (Chambers, Lee, Chun, & Miller, 2012). Thus, even though the source, kimchi, is specified, there still is room for interpretation. The word *kimchi* itself is just a noun that refers to a particular food item. Only by embedding it in a phrase with a perception verb, such as *It tastes like X*, does the noun's primary focus come to be a perceptual quality. Another limitation of source-based language is the reliance on common ground (see Levinson & Majid, 2014, p. 410): The identified source (in this case kimchi) needs to be known by both speaker and hearer. The description *It tastes like kimchi* does not help those who have never tasted kimchi.[2]

In the literature, discussions of source-based language most frequently come up in the context of taste and smell (e.g., Croijmans & Majid, 2016; Dubois, 2007). Lehrer (2009, p. 249) states that "most expressions for describing smell are based on an object that contains the odor." However, it has to be emphasized that source-based language comes up in the context of all senses.[3] Sound, for instance, is another domain where source-based language is quite frequent: When asking the question *What is that sound?*, speakers may in fact expect an answer that identifies the source of the sound, such as *It is a squirrel*, rather than an answer that merely describes the sound, such as *It is a cracking sound* (Huumo, 2010, pp. 56–57). David (1997, reported in Dubois, 2007, p. 179) found that for noises, 70.5% of the descriptions given by participants were source-based.

Source-based language may be particularly frequent in expert language, such as the language of wine and coffee experts (Croijmans & Majid, 2016; Lawless, 1984). This may have to do with the fact that source-based language, if the source is known, allows a lot of precision (Plümacher, 2007, p. 66). Alternatively, it may have to do with the fact that experts have more common ground with each other, often having undergone similar training. For example, expert descriptions of the quality of a wine may identify the chemical sources of a wine, such as *hydrogen sulfide* or *acetaldehyde* (e.g., Lehrer, 2009, p. 6). Such language use is entirely opaque to novices, who do not know the respective sources. Lee (2015) reports the results of a smelling test with speakers of the language Amis, an indigenous language of Taiwan from the Austronesian language. She finds that even though this language has a much larger vocabulary of abstract smell terms than English, speakers still predominantly talk about smells using source-based descriptions.

Source-based language in English does not exclusively rely on the use of *like*; other morphosyntactical patterns are possible. For instance, source terms can be used to modify existing color adjectives, such as in *tomato red*, *olive green*, *lemon yellow*, or *cobalt blue* (Dubois, 2007, p. 175; Graumann, 2007, p. 136; Plümacher, 2007, p. 62). New source-based sensory adjectives can furthermore be formed by derivational morphology, such as when adding the suffix *–y* to *onion*, forming *oniony* (Dubois, 2007, p. 175; Lehrer, 2009, p. 13), or to *salt*, forming *salty* (Ankerstein & Pereira, 2013, p. 313). Some source-based descriptions become conventionalized expressions that form a stable part of the English lexicon, as with adjectives such as *caramelized*, *citrusy*, and *nutty*. This, too, may more frequently occur in the domain of smell, with Lehrer (2009, p. 13) stating that "almost all words for smells are based on a noun denoting something with a distinctive smell." Examples of conventionalized source-based terms in the visual domain include *orange*, *peach*, and *salmon*. In some cases, sources may be entirely opaque. For instance, the English color words *crimson* and *carmine* both go back to the Arabic word *qirmiz*, which describes an insect, as well as the red dye that is produced from that insect. This is not known to most speakers of English, who are unaware that these adjectives ultimately have a source-based origin.

Other languages may have additional morphological means available to create source-based descriptions, such as reduplication. In the indigenous Austronesian language Amis, spoken on Taiwan, speakers can reduplicate a noun to indicate the smell corresponding to the noun's referent (Lee, 2015). For example, when the noun *ʔisi* 'urine' is reduplicated to form *hala-ʔisi-ʔisi* (*hala-* is a proclitic), the smell of urine is implied. The language isolate Yélî Dnye spoken on Rossel Island in Papua New Guinea similarly uses reduplicated nouns to convey color meanings (Levinson, 2000), such as when the word *taa* for 'red sparrow' is reduplicated (*taa-taa*) to convey redness, or when the word *kpaapî* 'white cockatoo' is reduplicated (*kpaapî-kpaapî*) to convey whiteness.

3.4. Describing perceptual qualities with arbitrary symbols

The final form of semiosis involves arbitrary symbols. This is what Clark (1996) paraphrases as "describing-as" (p. 187). In fact, most sensory words of English can be viewed as primarily belonging to the arbitrary encoding strategy. In the domain of sight, for instance, there are English words such as *bright*, *blue*, and *shiny*. There are also *rough* and *smooth* for touch, *loud* and *quiet* for sound, *sweet* and *sour* for taste, and *musky* and *fragrant* for smell. In each case, we can only know what these words mean by having learned the respective conventions.

However, once a word is known, sensory adjectives are both concise and precise. They directly relate to particular aspects of the sensory world, and in contrast to most source-based expressions, they do so using just one word. Moreover, conventionalized words are, by definition, an established part of the lexicon of a language. This means that they are understood by a larger portion of language users.[4]

Just as with the other strategies, communicating via arbitrary symbols "never works alone" (Clark, 1996, p. 187), with sensory adjectives often being part of composite signals. The deverbal adjectives *squealing* and *beeping*, for example, are clearly part arbitrary, part iconic. The descriptors *orange* and *citrusy* are clearly part arbitrary, part source-based. In real human interactions, any of the abstract words may be combined with vocal and manual gestures, as in the above-mentioned case of the relatively arbitrary word form *huge*, which can be phonetically lengthened (*huuuuuge*) and accompanied by co-speech gestures that signal size.

3.5. Technical language

Technical language can also be used to talk about sensory perceptions. Examples of technical language include describing a sound as having a frequency of 440 Hertz or a color as having a wavelength of 700 nanometers. Similarly, for anybody who knows hexadecimal codes, it is clear that *#FF0000* means 'red', and that *#0000FF* means 'blue'. Semiotically, technical language is largely based on abstract descriptors, including numbers. However, technical language is different from other abstract descriptors because it relies on identifying elements within a larger scientific system, such as the wavelength continuum for color and the frequency continuum for pitch. Likewise, hexadecimal codes are part of a larger system, and each hex code only makes sense with respect to that system.[5]

Moreover, technical language relies more heavily on cultural progress; it only works because humans have established knowledge about the physical characteristics of perception.

While a lot of technical language requires expert knowledge (see Porcello, 2004), some technical vocabularies have become common currency in the general population. English speakers can tell each other about time with reference to temporal measurement units such as *seconds*, *minutes*, *hours*, and *days*; and they can tell each other about distances by using spatial measurement units such as *meters*, *feet*, and *inches*. Crucially, just as is the case with the above-mentioned examples of visual wavelengths, acoustic frequencies, and hexadecimal codes, these words only make sense in relation to the entire system. To understand the expression *The trip took 30 minutes*, one needs to know the underlying scale of how time is counted in hours and minutes. Technical descriptions are always relative to a whole system.

Another feature common to technical language is that the used systems often have language-external manifestations, such as visual representations of color spectrums or tables of hexadecimal codes. People who are familiar with such representations can refer to perceptual qualities by metaphorically talking about spatial locations or movements within such systems. This is very common in music (Pérez-Sobrino & Julich, 2014): For example, the circle of fifths is a (metaphorical) spatialization of a musical system that captures the relationship among the 12 tones commonly used in Western music, as well as their role in keys (major and minor). Given this system, a musical expert can say such things as *a series of chromatically <u>ascending</u> tritones*, or *Lohengrin's A major (..) <u>stands the furthest away</u> in the circle of fifths* (Pérez-Sobrino & Julich, 2014, p. 305). Here, spatial language is used to refer to positions and movements within the established technical structure.

Interestingly, taste and smell appear to have no scientific representational systems that are grounded directly in physical facts (see Dubois, 2007, p. 171). Taste and smell arise from complex interactions of many molecules, with limited success in mapping the chemical structures of odorants to the psychological characteristics of odors (compare Agapakis & Tolaas, 2012). There are expert representational systems, such as the flavor wheels that are commonly used in wine discourse (e.g., Gawel, Oberholster, & Francis, 2000; Noble et al., 1987; see also Lehrer, 2009, Chapter 4), but in contrast to such systems as the circle of fifths in music or the wavelength description of color, these systems are largely intended to standardize description; they are less grounded in physical facts.[6]

In part as a result of the lack of representational formats that allow the detailed characterization of smell, Sissel Tolaas developed the international language Nasalo (akin to Esperanto), consisting of words to communicate smells, such as *dusbi* for the smell of 'dusty brick' and *isjfe* for the 'smell of cut grass' (Agapakis & Tolaas, 2012, pp. 570–571). In fact, sensory vocabularies (albeit not all as creative as the one of Tolaas) are continuously being developed in order to meet the demands of the food industry. Almost every issue of the journal *Journal of Sensory Studies* contains a new lexicon for a specialized sensory domain, such as lexicons for the description of nail polish (Sun, Koppel, & Chambers, 2014), artisan goat cheese (Talavera & Chambers, 2016), or kimchi (Chambers et al., 2012). These expert systems are designed to overcome some of the limitations of the everyday English sensory vocabulary. They are intended to establish common ground within a particular community of experts, as well as to standardize descriptions within this community (see Diederich, 2015).

3.6. Metaphor

Metaphor is a very versatile strategy for conveying perceptual meaning. Words such as *sweet* and *smooth* appear to be primarily about taste and touch perceptions when seen in isolation. However, both words can easily be used to describe sensations more strongly related to other senses, such as when speaking of *sweet melodies* and *smooth tastes*. These expressions have been called "synesthetic metaphors" and are associated with a large literature (e.g., Shen, 1997; Strik Lievers, 2015; Ullmann, 1959) that will be the topic of Chapters 6 through 9.

Sensory words are constantly used in a crossmodal fashion—that is, outside of their core sensory domain. In her discussion of wine language, Lehrer (1978, p. 106) says that "many of the terms used as wine descriptions involve extensions of meaning rather than standard senses" (see also Suárez-Toste, 2013). All word meanings, including sensory word meanings, are flexible and can be modulated by context as well as creatively used in entirely novel domains.

Metaphor is one of the primary strategies to enrich vocabularies (Dirven, 1985; Ortony, 1975), including sensory vocabularies. For example, a particular sound may be difficult to describe using sound words alone, so touch words can be used instead, as in *sharp sound* and *abrasive sound* (e.g., Day, 1996; Ullmann, 1959; Williams, 1976). The language of sound (Barten, 1998; Porcello, 2004) and music is particularly replete with metaphors, including spatial metaphors such as *high pitch* and *low pitch* (Dolscheid, Shayan, Majid, & Casasanto, 2013). According to Pérez-Sobrino and Julich (2014), about 29% of all words used in academic discourse on music is metaphorical, which is higher than in other domains (cf. Steen et al., 2010). Over the last 200 years, the frequency of sensory metaphors has increased (Akpinar & Berger, 2015). Because metaphor is such a dominant strategy in talking about sensory perception, four entire chapters will be devoted to this topic (Chapters 6–9).

3.7. Summary

Humans are expert meaning-makers that have a wide array of semiotic tools available to them, each of which is capable of conveying perceptual meaning. I will conclude by highlighting some of the dimensions that crosscut the different semiotic strategies and can be used to compare and contrast them.

One such dimension is whether a perceptual quality is identified directly or indirectly. Conventionalized words can directly relate to perceptual qualities. Iconicity may be even more direct because it connects to a sensory meaning via resemblance. Yet, all other semiotic strategies are indirect. This is the case with source-based language, where speakers do not specify a perceptual quality directly, but they instead refer to the conditions that elicited a particular percept. This is also the case with technical language, where speakers refer to elements within a scientific system. And this is also the case with metaphor, where speakers borrow perceptual language from another domain to talk about a particular sensory concept.

The directness of abstract words such as *red* and the directness of iconic forms such as *squealing* are, however, of a fundamentally different kind. Abstract words strip away from the particularities of experience and identify a percept as an instance of a general category. On the other hand, iconic forms—especially if they are prosodically modulated—stay closer to the particularities of experience. This is one of the reasons why it has been argued that iconicity and abstraction are incompatible (Lupyan & Winter, 2018).

Another theme that ran through this chapter was the issue of common ground. Some semiotic encoding strategies are more reliant on knowledge that is shared between the speaker and hearer. This is the case particularly for source-based language, and also for technical language. Conventionalized sensory words and iconicity require less common ground. Thus, there is a correspondence between the directness with which an encoding strategy relates to perception and its reliance on common ground: More indirect strategies generally also require more common ground.

Another dimension that crosscuts the different semiotic strategies is whether a particular form of sensory expression is encoded in the lexicon or not. Abstract words always are encoded in the lexicon; in fact, their conventionalized nature is definitional. On the other hand, iconicity may or may not be encoded in the lexicon (e.g., phonetic iconicity versus phonological iconicity, as in *huuuuuge* versus *squealing*), and the same applies to source-based language, which can be generated on the fly (*It tastes like kimchi*) or which forms a stable part of the English lexicon (the color word *orange*).

Finally, it should be emphasized again that any analysis of sensory language has to keep the full system of semiotic strategies in mind. This is for at least two reasons: First, as was mentioned above, the different semiotic strategies are often employed in tandem (Clark's composite signals). Second, speakers always select one encoding strategy in relation to all others. That is, speakers have a *choice* about which tool of their semiotic toolkit they want to employ. The trade-offs between different semiotic strategies are discussed in following chapter.

---

[1] The only empirical data for shape iconicity in word forms is that words for round objects across the world's languages often contain the phoneme /r/ (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016).

[2] Another way to analyze source-based language linguistically is to view it as a form of metonymy (Cacciari, 2008, p. 426), which will be discussed in more detail in Chapter 7. A comparison to an expression such as *The White House has pardoned former sheriff Joe Arpaio* is

useful. This expression involves a PLACE STANDS FOR INSTITUTION metonymy (names of common metonymies and metaphors are often presented in capitalized letters), where the expression *White House* (PLACE) is used to reference the related concept of the US government (INSTITUTION). This is structurally similar to the case of source-based language, where a perceptual quality is identified via reference to another concept, its source. Moreover, just like with source-based language, metonymy relies heavily on shared cultural knowledge. In the aforementioned case of the PLACE STANDS FOR INSTITUTION metonymy, the listener has to know that the White House is where the President of the United States resides and that the White House is where the seat of the government is.

[3] Fainsilber and Ortony (1987) discuss source-based language in the context of emotions, such as when a speaker characterizes a particular emotional state, such as embarrassment, by describing a particularly embarrassing episode. They say that "in such a case, the literal description would not describe the *quality* of the subjective state itself but would merely identify its eliciting conditions" (p. 241).

[4] As clarified by Plümacher and Holz (2007, p. 5), some abstract language about sensory perceptions is subject to additional constraints, such as the word *blonde*, which is a color term that can only be used to talk about hair. Moreover, some abstract language about sensory perceptions may have additional meanings, such as the expression *green tomato*, which may not only describe the color of a tomato, but also the fact that it is unripe.

[5] The arbitrariness of technical language becomes clear when we consider the fact that such dimensions as wavelength and acoustic frequency could have been defined via another numerical scale. Hexadecimal codes are particularly arbitrary and in fact, different color representation systems exist, such as RGB values in the 0–255 range.

[6] There are other differences between flavor wheels and such systems as the circle of fifths. The circle of fifths is continuous. Flavor wheels give the impression of a continuous scale, but they are in fact categorical and hierarchical (e.g., *cut grass* and *dill* may be co-hyponyms of the hyperonym *fresh herbs*, which itself is a hyponym of *vegetal*); see discussion in Lehrer (2009, Chapter 4).

Perhaps as a result of their categorical nature, there is no work of which I am aware that has described speakers using metaphorical movement language within these systems, in contrast to the circle of fifths (Pérez-Sobrino & Julich, 2014).

Chapter 4. Ineffability
4.1. Introduction

Consider the following two examples that contain perceptual descriptions of kimchi taken from the Corpus of Contemporary American English (Davies, 2008).

(1) *No Korean meal is complete without kimchi, the pungent, piquant cabbage-based side dish.*
(Vegetarian Times)

(2) *But pause before calling kimchi the Korean sauerkraut.... It's not just the red pepper, garlic and ginger that set kimchi apart from sauerkraut.*
(Austin American Statesman)

Have you ever eaten kimchi? If yes, ask yourself: How would *you* describe the experience of eating kimchi to a friend who has never tasted it? If you have never eaten kimchi, ask yourself: How do you think kimchi tastes, given the descriptions listed above? How would it smell like? In fact, one may ask: Is there any linguistic description you or another English speaker could possibly give that would adequately capture the full delight of eating kimchi?

The linguistic descriptions above provide a good starting point for getting an idea about kimchi, but none of them capture the richness the dish's qualities. Even the standardized sensory lexicon of kimchi flavors (Chambers et al., 2012), detailing such characteristics as "crispness" and "tongue tingle," is not able to convey all aspects of the multisensory kimchi-eating experience. In talking about how limited language is when it comes to expressing sensory perceptions, Moore and Carling (1988, p. 110) state that "if someone wants to know what caviar or kumquats taste like, better they taste some for themselves."

This chapter goes into more detail about the difficulty of translating sensory perception into words. I will move from definitions of ineffability (Chapter 4.2) to discussing what exactly it is that is ineffable or not (Chapter 4.3). Then, I will discuss different explanatory accounts of ineffability (Chapter 4.4). I will conclude by revisiting the topic of ineffability from the perspective of the semiotic toolkit introduced in the last chapter.

4.2. Ineffability and related notions

Levinson and Majid (2014) define ineffability as "the difficulty or impossibility of putting certain experiences into words" (p. 408). Levinson and Majid's notion of ineffability is really about the existence or absence of dedicated linguistic material for a given perceptual quality (p. 411).

Levinson and Majid (2014) distinguish between weak and strong ineffability. Weak ineffability is language-specific; strong ineffability is absolute. Weak ineffability means that something can be expressed in one language, but not in another. For example, the Amazonian language Pirahã lacks some of the basic color terms that English has, such as *red*, *green*, and *blue* (Everett, 2005). The fact that English has words that Pirahã lacks shows that these perceptual concepts are principally codable. Conversely, the Maniq language spoken by hunter-gatherers in Thailand (Wnuk & Majid, 2014) has many more smell terms than English, including such words as *caŋɛs* for the smell of monkey hair and burnt animal hair. This means that this sensation is principally expressible; it just so happens that English lacks the corresponding word. On the other hand, strong ineffability means that something cannot possibly be said in any language. The only way to establish whether something is weakly or strongly ineffable is via crosslinguistic comparison.

Ineffability needs to be further distinguished from “efficient codability,” “communicative accuracy,” and “conveyability / indirect indication” (Levinson & Majid, 2014). Efficient codability is a psycholinguistic measure of the relative ease of expressing certain percepts. Brown and Lenneberg (1954) think of efficient codability as a multivariate concept that includes several indicator variables. According to them, sensory perceptions are more efficiently codable if speakers give shorter descriptions (i.e., less words are needed), if the words they use are shorter (i.e., no long words are needed), and if the words can be named more quickly (i.e., highly accessible and easily pronounceable words are used). For instance, speakers of the Austroasiatic language Jahai are faster and more succinct in describing odors than Dutch speakers (Majid et al., 2018), suggesting that smell is more efficiently codable in Jahai than in Dutch; that is, the Jahai people find odors easier to talk about than Dutch speakers. Brown and Lenneberg’s (1954) notion of codability also includes whether speakers agree on the same labels and whether a given speaker uses a label consistently across time. Majid and Burenhult (2014) show that in a smell labeling task, Jahai speakers agree more with each other than English speakers.

An extension of Brown and Lenneberg’s (1954) notion of efficient codability is the notion of communicative accuracy, which measures whether the labels of one group of participants allows another group of participants to correctly identify the intended sensory stimulus (see discussion in Lucy, 1992, Chapter 5).

Finally, conveyability refers to the fact that perceptual concepts having no dedicated words associated with them can always be paraphrased—for example, by using the source-based strategy (e.g., *It tastes like kimchi*; compare Cacciari, 2008, p. 426). Levinson and Majid (2014) discuss conveyability in the context of Searle’s Principle of Expressibility, according to which anything that can be thought can also be expressed (Searle, 1969). This principle is connected to the generative capacity of language. In any language, it is possible to say a sentence that has never been said before.

4.3. Ineffability of what?

We need to ask the question: What exactly is it that is ineffable or not? For example, is it possible to meaningfully speak of an entire sense, such as sight or smell, as more or less ineffable? Or is ineffability a feature of particular perceptual qualities within a given sensory modality? These questions will be discussed in (Chapter 4.3.1), followed by a discussion of perceptual characteristics that are shared between sensory modalities, so-called "common sensibles" (Chapter 4.3.2). Finally, I discuss ineffability of other aspects of sensory perception, in particular:

1. The ineffability of subjective experience (Chapter 4.3.3)
2. The ineffability of fine perceptual detail (Chapter 4.3.4)
3. The ineffability of the multisensoriality (Chapter 4.3.5)

4.3.1. Differential ineffability of the senses

It is generally thought that sight is the most codable sensory modality, with linguists having suggested that there is more lexical differentiation for visual concepts in the world's languages than for the other senses (i.e., more distinct visual word types; Buck, 1949, Chapter 15; Viberg, 1983). Levinson and Majid (2014) say that "in English at least, it seems generally easier to linguistically code colors than (non-musical) sounds, sounds than tastes, tastes than smells" (p. 415).[1] Slobin (1971) already noted that there is "an inadequate vocabulary for expressing sensations of the proximity senses" (p. 108), which includes touch, taste, and smell—in contrast to the so-called "distal" senses of sight and sound. The evidence from type frequencies suggests that indeed visual concepts are overall more effable in English, and that taste and particularly smell are relatively ineffable (see Chapter 12). Vision is also overall more efficiently codable, as evidenced by the fact that visual concepts have higher token frequencies (e.g., San Roque et al., 2015; Viberg, 1993; Winter et al., 2018) and are processed more quickly (Connell & Lynott, 2014).

Compared to vision, smell in particular has been argued to be relatively ineffable. Yeshurun and Sobel (2010) mention that if one were to subject their friends to a blind smelling test with food and beverage items from their own fridge, they would be incapable of identifying even those items they consume on a daily basis. While it is often easy for speakers to recognize particular smells, research has shown that labeling smells is difficult (Cain, 1979; Croijmans & Majid, 2016; de Wijk & Cain, 1994; Engen & Ross, 1973; Huisman & Majid, 2018; Yeshurun & Sobel, 2010). This is one example of the "persistent challenges" in "mapping odors to names," which is why researchers have called smell a "muted sense" (Olofsson & Gottfried, 2015, p. 319). As stated by Lorig (1999, p. 392), "odor and language do not seem to work well together." Smell is generally seen as the most ineffable sensory modality.

However, Levinson and Majid (2014) also stress that ineffability does not apply to entire sensory modalities so as much as to particular perceptual qualities within a given modality. They say that "within the visual domain, object shapes seem easier to verbalize than faces..., while within the tactile modality textures seem easier than pain" (p. 415; for the difficulty of talking about faces, see Moore & Carling, 1988, Chapters 13–14). Within the auditory modality, it may be relatively easy to talk about loudness (*loud, quiet*) and tempo (say, of a musical piece; *fast, slow*), compared to talking about timbre and spectral characteristics, for which speakers frequently use touch-related vocabulary (*rough sound*, *harsh sound*, *abrasive sound*). Within the modality of taste, there are dedicated words for basic tastes (*sweet*, *sour*, *bitter*, and *salty*) but source-based descriptors for many other gustatory experiences (*oniony*, *nutty*, *vinegary*). Also, it is the case that although the detailed descriptive characteristics of smells are not encoded in the English lexicon, perceived pleasantness is, as reflected in such words as *aromatic* and *pungent*. Thus, when we say that sight is overall more codable in English, what we actually mean is that sight has more distinct perceptual qualities that are codable.

4.3.2. Proper and common sensibles

When talking about what particular perceptual qualities are linguistically codable or not, it is important to distinguish between "proper sensibles" and "common sensibles," a distinction which goes all the way back to Aristotle (Sorabji, 1971). Proper sensibles can only be perceived through one sensory modality; for instance, color is a proper sensible of vision. Common sensibles refer to properties that "can make themselves known through several sensory channels" (Marks, 1978, p. 12), such as distance, speed of movement, shape, or duration (Levinson & Majid, 2014, p. 412; see also discussion in Ronga et al. 2012, pp. 149–150). All spatial properties can be perceived through sight, touch, and sound.

A special common sensible is magnitude. Levinson and Majid (2014) say that "lights, sounds, smells, tactile pressures, tastes, pains, emotions can all have low or high intensities" (p. 413). Walsh (2003) and Bueti and Walsh (2009) propose what they call ATOM (A Theory of Magnitude), according to which all kinds of sensory magnitudes (including numerical quantities) are processed by a crossmodal magnitude system (see also Winter, Marghetis, & Matlock, 2015).

According to this theory, magnitude as it is experienced via different modalities is actually the same underlying cognitive quality.

To say that something is a common sensible does not mean that all senses have equal access to the given perceptual quality. In some tasks, two modalities may be nearly equivalent, as when the distance of rods is judged by touch and by sight (Teghtsoonian & Teghtsoonian, 1965). However, when the perception of a common sensible is tested on two modalities together, asymmetries may emerge. In a series of classic experiments, Rock and Victor (1964) and Hay and Pick (1966) showed that in shape perception, sight influences touch more than the other way around. Another asymmetry characterizes touch and sound. The perceptual quality of roughness is a common sensible to both touch and sound, but sound has little effect when surfaces can also be felt (Lederman, 1979). These asymmetries may carry over to language. For example, even though *rough* may also describe sounds, native speakers of English more strongly associate this word with touch (Lynott & Connell, 2009). Similarly, even though *high* and *low* describe spatial characteristics that are accessible to multiple senses, native speakers think of these words as primarily visual (Lynott & Connell, 2009).

As a result of their intrinsic multisensoriality, common sensibles are less confined to the ineffability of any one sensory modality. Because of this, common sensibles are generally easy to express in language. For instance, length can be expressed with words such as *long* and *short*, but also through iconic gestures (as when extending the hands slowly to indicate a long duration), vocal gestures (as when lengthening the word *looooong*), and through specialized spatial language, such as measurement units (*50 centimeters, 2 meters*). Essentially, all strategies discussed in Chapter 3 can be applied to most common sensibles.

The existence of common sensibles makes the measurement of ineffability difficult.[2] When counting how many word types a sensory modality has, are words such as *large* and *rectangular* treated as belonging to sight, to touch, or perhaps to no sense at all? Ronga et al. (2012, p. 159, Footnote 24) discusses how such "dimension words" are treated by Popova (2005) as touch-related, and by Williams (1976) as sight-related. Such mismatches are to be expected for common sensibles. In general, because common sensibles can be accessed through multiple sensory modalities, it is not clear what access modality is primarily encoded in a linguistic form. The perceptual ratings described in more detail in Part II of this book (specifically, Chapter 11) allow dealing with this multisensoriality in a principled fashion. For example, to look at whether a certain sensory modality is associated with more or less words, it is perhaps best to use the least multisensorial part of the vocabulary (as is done in Chapter 17). More generally, differences in classification of common sensibles (such as highlighted by Ronga et al., 2012) can be used by using standardized word lists.

In the following sections, I move to a discussion of different types of ineffabilities.

4.3.3. Ineffability of subjective experience

Each and every one of us perceives the world in a slightly different fashion. This is perhaps most apparent in the case of the so-called "chemical senses," taste and smell. For instance, Miller and Reedy (1990) report striking differences in how many taste buds people have on their tongue. In their small-scale study of only 16 participants, the density of fungiform papillae (a particular type of taste bud) ranged from about 22 per square centimeter for some individuals to 74 per square centimeter for others, which also resulted in differences in subjective intensity thresholds for particular taste sensations. Taste and smell also vary as a function of age, gender, culture, and many other factors. Spence and Piqueras-

Fiszman (2014) review many individual differences in flavor perception and conclude that "we all live in very different taste worlds" (p. 199).

Similar things can be said about any other sense. For instance, there is great variety in the number of photoreceptors people have within their eyes (Curcio, Sloan Jr., Packer, Hendrickson, & Kalina, 1987), and there are several forms of color blindness, some of which are quite common in the general population (e.g., red-green color blindness affects about 4–8% of men, depending on the population studied; Birch, 2012).

The linguistic system ignores many of these individual differences. Shepard and Cooper (1992) showed that when participants with varying forms of visual impairment are asked to perform similarity judgments on the color words *red*, *orange*, *gold*, *yellow*, *green*, *turquoise*, *blue*, *violet*, and *purple*, the resulting similarity space looks just like the similarity space that is created when fully sighted people make those judgments. Only when judgments are made on actual color stimuli do strong differences become apparent. Similarly, Landau and Gleitman (1985) show that even congenitally blind children can acquire the meaning of color words. This demonstrates how many differences between individuals may be masked by language.

Moore and Carling (1988) state that "the meanings of words are necessarily vague and necessarily variable in the minds of different individuals" (p. 44) and that "language is based on a tacit communal conspiracy whereby we assume, individually, that we mean broadly the same by the same words" (p. 59). The fact that public language use is targeted at shared meanings necessitates disregarding genuinely subjective experience in linguistic encoding. Levinson and Majid (2014) say that "for a community to share the meaning of a word, they must all have direct or indirect access to the referent" (p. 420), which is not the case with those aspects of experience that are truly different between different individuals. Holz (2007, p. 193) talks about the "apparently unclosable gap between the individuality of perception and the conventional, hence social and communicative character of language." Thus, truly subjective sensations—what philosophers call qualia—may be truly ineffable in an absolute fashion.

4.3.4. The ineffability of fine perceptual detail

Even assuming that every speaker associates exactly the same meanings with perceptual words, there would still be a massive gap between what can be perceived and what is linguistically encoded. A case in point is the fact that humans can distinguish between millions of different colors (some estimates go up to 10 million distinct colors; Judd & Wyszecki, 1975), but languages generally have only very few basic color terms (Berlin & Kay, 1969; see discussion in Fahle, 2007; Graumann, 2007, p. 129; Wyler, 2007). Humans can similarly distinguish thousands of different smells (see Agapakis & Tolaas, 2012; Yeshurun & Sobel, 2010), but at least in English, there are very few smell words. In any sensory modality, the number of possible perceptual experiences vastly outnumbers the size of any semiotic toolkit (cf. Staniewski, 2017). This numerical mismatch means that linguistic labels necessarily have to be fuzzy; that is, fine-grained distinctions are hidden when speakers use sensory words (see Fahle, 2007). The word *red*, for instance, encompasses a whole range of wavelengths that speakers do not distinguish when they use the word. Technical language makes it possible to communicate fine perceptual detail more precisely (e.g., by specifying the exact wavelength of a color), but such technical language is generally not available to most language users. Wyler (2007) states that language as a social system does not "comply" with the "physiological precision" of actual perception (p. 116). Altogether, it looks as if the linguistic system ignores a considerable amount of fine perceptual detail, no matter what sensory modality is involved.

4.3.5. Ineffability of multisensory experience

Finally, one has to consider the fact that perception always engages all of the senses in concert (Spence, 2011; Spence & Bayne, 2015). Spence (2012, p. 37) says that “most of our everyday experiences...are multisensory.” For example, let us consider the intense multisensoriality that goes into such a simple event as eating kimchi. This experience involves the salty and spicy mélange of pepper and garlic notes that excite the taste buds, on top of the fermented smell, the tingly mouthfeel and the crunchy chewing sound (compare Chambers et al., 2012). However, conveying this experience forces the use of decoupled sensory adjectives such as *salty* and *crunchy*, as is the case with my description above. The compression inherent in these words, each one singling out one aspect of experience, means that the simultaneity of the multisensory taste experience cannot be conveyed. The fact that flavor perception is perhaps “one of the most multisensory of our everyday experiences” (Spence & Piqueras-Fiszman, 2014, p. 183; see also Spence, 2013, 2015) is impossible to communicate via the sequential format of language.

4.3.6. Why ineffabilities are necessary

It is possible to construe the ineffabilities discussed so far as disadvantageous. This sentiment is expressed by Cacciari (2008, p. 425), who says that “the informational richness of perceptual experiences is hardly rendered by linguistic expressions.” From this perspective, all sensory language is just a pale reflection of our sensory worlds; even the best writers or poets will never be able to communicate all of that which is perceivable.

However, it is possible to view the loss of subjective experience and fine perceptual detail, as well as the compression of multisensoriality, as advantageous traits. Moore and Carling (1988) say that "in using language we necessarily reduce and group and select" (p. 19). Selection means to highlight particular aspects of sensory experience, those aspects that are relevant to speakers and hearers within a given conversation. Grouping involves saying that a particular experience is a member of a more abstract category of experiences (Moore & Carling, 1988, p. 20). For example, in labeling a taste experience as *sour*, a speaker establishes similarity with other experiences, thus moving away from a particular experience to the general characteristics of sourness (Lupyan & Clark, 2015, p. 283). Abstraction furthermore allows for intersubjectivity and sharability: All native speakers of English have an understanding of the word *sour* that is at least partially shared, precisely because speakers understand that the word *sour* is not about a particular perception that happens in a specific moment in time but about a general characteristic.

From this perspective, the sensory adjectives which will be the focus of Part II of this book, such as *sour* and *greasy*, can be seen as what I call "compression devices." That is, sensory adjectives single out particular perceptual qualities at the expense of other qualities (Gärdenfors, 2014, Chapter 7; Paradis, 2005). Givón (2001, p. 53) states that "the most prototypical adjectives are single-feature concepts, abstracted out of more complex bundles of experience." In many ways, reducing the complexity of the perceptual world is the true purpose of sensory adjectives, which allows for abstraction, generalization, and intersubjectivity. However, the flipside of these advantages is that subjective experience, fine perceptual detail, and multisensoriality are truly ineffable.

In fact, Levinson and Majid (2014) implicitly link their notion of ineffability to the type of compression discussed here. According to them, saying that color is "codable" means that there are "words which code only the descriptive, abstract property of color and not other properties such as reflectance, texture, shape etc." (p. 411). However, it should be emphasized that all of the semiotic strategies discussed in Chapter 3, including iconicity and source-based language, single out particular aspects of perceptual experience. All semiotic strategies involve compression.

It furthermore must be pointed out that not all of the compression of multisensoriality happens at the linguistic stage. Some compression is prelinguistic and happens as a result of two processes: first, selective attention, which is the cognitive capacity to single out particular aspects of the perceptual world (Spence, 2010). Thus, our mind already filters out a lot of the perceptual morass humans are exposed to before linguistic compression even begins.

The second prelinguistic form of compression results from the folk concepts of the senses. Speakers are generally unaware of the intense multisensoriality of their perceptual worlds; for example, most crossmodal correspondences (such as associations between brightness and pitch) are not consciously accessible (Deroy & Spence, 2013). Spence (2012, p. 37) reminds us that "introspection often tells us that we see color only with our eyes, that we feel softness exclusively with our fingertips, and that we taste...only with our mouths." Humans can only communicate about that what is accessible to their conceptual world. Since we live under the illusion that our senses are separate (see also Auvray & Spence, 2008; Spence, Smith, & Auvray, 2015), it is this illusion that we communicate to others and becomes conventionalized in the lexicon. Pink (2011, p. 266) speaks of "a rather less culturally structured flow of neurological information" that "becomes differentiated into categories that we call the senses." She then says "we tend to communicate linguistically about our embodied and sensory perception in terms of sensory categories," but warns us that "because one category is never enough to express exactly what we have actually experienced, the illusion of 'separate' senses...is maintained."

Thus, to conclude this section, the compression of the perceptual world happens in part due to the constraints of the linguistic system itself, e.g., due to the specialization of sensory words. However, in addition, there is a prelinguistic stage of compression (compare Miller & Johnson-Laird, 1976, Chapters 1–2), stemming from selective attention as well as from folk categorizations of the sensory world.

4.4. Explanations of the ineffability of the senses

From now on, the focus will shift to what explains the "differential ineffability of the senses" discussed above. In particular, what explains the fact that certain perceptual qualities are associated with more word types than others?

Levinson and Majid (2014, pp. 417–421) discuss two broad strands of explanations that try to account for such ineffability: "Cognitive-architectural" explanations, which focus on perception, and "limits of language" explanations, which focus on language. I will add to this a third explanatory approach, which is the one of "communicative need." The three types of explanations cannot neatly be separated, but they will help to structure the following discussion. To elucidate the different explanations, the discussion will focus on sight and smell, because these are presumed to be the respective end points of ineffability in English, with sight being the most and smell being the least effable.

1. cognitive-architectural explanations (Chapter 4.4.1)
2. limits of language explanations (Chapter 4.4.2)

3. communicative need explanations (Chapter 4.4.3)

4.4.1. Cognitive-architectural explanations

The defining feature of cognitive-architectural explanations is that they resort to extralinguistic facts, such as facts about perception or the brain. There are many possible cognitive-architectural facts that can be used to explain the differential ineffability of sight and smell. Lorig (1999) claims that both language and odors share complex temporal signatures, which means that they are competing for neural resources. Yeshurun and Sobel (2010) argue that the mental representation of odors is in terms of their emotional effects on the perceiver, ultimately yielding a representational format that is not very amenable to linguistic encoding. Olofsson and Gottfried (2015) argue that smell lacks the dedicated naming pathways that vision has, such as the ventral visual pathway.

These explanations focus on the neural and cognitive mechanisms that link perception and language. Other cognitive-architectural explanations look at asymmetries between the senses more generally, with the idea that senses that are perceptually dominant will also be linguistically dominant.

Much research suggests that vision dominates the other senses in a whole range of perceptual tasks, an idea called "visual dominance" (Stokes & Biggs, 2015). As was discussed, vision dominates touch in shape perception (Hay & Pick, 1966; Rock & Victor, 1964). In the so-called "ventriloquist effect" (Pick, Warren, & Hay, 1969; Welch & Warren, 1980), vision dominates sound in location perception. Vision also has profound effects on taste: Morrot, Brochet, and Dubourdieu (2001) showed that dying white wine red led oenology undergraduates to use red wine terminology. Similarly, Hidaka and Shimoda (2014) showed that the coloring of a sweet solution affects sweetness judgments (see also Shermer & Levitan, 2014). More broadly construed, any advantage of vision over the other modalities is part of visual dominance, including attentional advantages (Spence, Nicholls, & Driver, 2001; Turatto, Galfano, Bridgemann, & Umiltà, 2004) and the fact that vision arguably takes up more cortical space (Drury et al., 1996).

Perhaps vision is also most advantaged when it comes to the perception of space, which helps us to navigate the world. Taste and smell appear to be inherently non-spatial (compare Miller & Johnson-Laird, 1976, p. 27). Compared to touch and sound, sight makes more spatial information available in a simultaneous fashion, including complex relations between objects (see Stokes & Biggs, 2015). All of these are language-external factors that may bias language toward the visual modality, at the relative expense of the other senses.

There also is evidence suggesting that when performing conscious mental imagery, people report being able to see what they imagine, but not to smell what they imagine (Arshamian & Larsson, 2014; Brower, 1947). Kosslyn and colleagues (Kosslyn, Seger, Pani, & Hillger, 1990) asked participants to jot down their experiences of mental imagery every hour of the day, finding that visual mental images occurred very frequently, especially when compared to olfactory images. This suggests that our inner world of mental imagery is biased toward the visual.

As a final avenue toward explaining ineffability with recourse to cognitive-architectural factors, consider the fact that smell is intensely malleable and variable—more so than sight. Agapakis and Tolaas (2012) say that "much of the study of olfaction shows that even the simplest judgments of odor quality are highly context-dependent, changing and shifting depending on molecular, biological, emotional and social contexts" (p. 569). For example, participants that initially rated a sweet smell as positive perceived it to be less pleasant after being injected with glucose (Cabanac, Pruvost, & Fantino, 1973; see also Cabanac, 1971). The chemical substance indole was reported to smell more pleasant when it was labeled *countryside farm* as opposed to *human feces* (Djordjevic et al., 2008). Spence and Piqueras-Fiszman (2014) detail the many ways through which the actual sensory qualities of a food dish (including both taste and smell) can be influenced via the name of a dish and via its visual presentation. Even music playing in the background of a dining experience can change flavor perception (Crisinel et al., 2012). There also are strong individual differences in what odors are perceived as pleasant (Hermans & Baeyens, 2002; Köster, 2002). For instance, some people perceive the smell of skunk (the animal) as pleasant, a smell which others abhor (Herz, 2002, p. 161).

The intense malleability and variability of smell may impede linguistic encoding because there are no stable perceptual targets to encode, and because linguistic communication, being about sharable phenomena, disprefers that what is subjectively variable. In comparison, sight appears to make more stable objects available to our consciousness, which we can use to coordinate with others via joint attention and thus use to establish common ground (e.g., San Roque et al., 2015, p. 50). This stability, compared to the variability of smell, may explain part of why there are so many visual words compared to smell words.

4.4.2. Limits of language explanations

According to Levinson and Majid (2014), explanations of ineffability grounded in the limits of language start with the assumption that particular sensations are accessible but not coded for linguistic reasons. They say that "arguments about the limits of language are likely to focus on the general intrinsic limitations imposed by the design features of a natural language" (p. 419). The digital nature of language discussed above (that sensory words are compression devices for multisensory experience) is one factor that results in the ineffability of multisensoriality. Vocabulary size may be an additional limiting factor, which restricts the range of sensory words (leading to ineffability of fine perceptual detail, discussed above). Limits of language explanations, however, are better for explaining such ineffabilities as the ineffability of multisensoriality and the ineffability of fine perceptual detail. They are not intrinsically well-suited to explain the differential ineffability of the senses. For example, vocabulary size limitations alone cannot tell us why it is particularly smell that is less lexically differentiated in English and not any other sense.

4.4.3. Communicative need explanations

Perhaps the simplest explanation of the visual dominance observed in English and other languages is that the speakers of these languages have a higher communicative need to talk about visual concepts as opposed to other concepts, particularly smell-related ones. Plümacher and Holz (2007, p. 2), for instance, talk about the "less developed cultural need to reflect odors" in language. According to Holz (2007), "we are very seldom confronted with the question of describing a smell by linguistic means" (p. 186).

Communicative need has the advantage of accounting for two established facts about visual concepts in one swoop: first, the fact that visual concepts have higher type frequencies—that is, higher lexical differentiation (Chapter 12; Buck, 1949, Chapter 15; Levinson & Majid, 2014; Strik Lievers & Winter, 2018; Viberg, 1983; Winter et al., 2018)—and second, the fact that visual concepts have higher token frequencies (Chapter 15; Viberg, 1993; San Roque et al., 2015; Winter et al., 2018). A correlation between type and token frequencies is exactly what is expected if communicative need is a driving force, since it shows precisely that the domains more frequently verbalized also afford more expressivity (Regier, Carstensen, & Kemp, 2016; Winter et al., 2018). However, this still leaves open the question: What explains the increased communicative need to talk about visual concepts in the first place?

To answer this question, one needs to distinguish the importance of a sense in perception from the desire of humans to talk about a sense. These two are at least logically independent; for example, is smell not important as a sense, or do humans not talk about smell despite its perceptual importance?

Some authors have claimed that humans have a diminished sense of smell compared to other animals. Gilad and colleagues compared the human genome to other primates, finding that only human primates showed an increased loss of olfactory receptor genes, which constitute the basis for the sense of smell (Gilad, Man, Pääbo, & Lancet, 2003). Moreover, the fact that this loss coincides with the acquisition of full trichromatic vision in the primate lineage (Gilad, Wiebe, Przeworski, Lancet, & Pääbo, 2004) suggests that humans have traded olfactory ability for visual ability.

On the other hand, several authors have pointed out that smell is, in fact, immensely important. For instance, smells affect mate choice (Havlicek & Roberts, 2009) and help us to avoid poisonous foods, both of which are vital to evolutionary success. Lorig (1999) uses the multi-billion dollar fragrance industry to argue for the importance of smells, saying that "the companies in this business sell products to scent shampoo, deodorants, tissues, soaps of all types, hand creams, leather products, toys, air fresheners, cleaning products, and many other commodities" (p. 391). According to Lorig, such high volume of spending would be unlikely if smell was, in fact, unimportant. Similarly, Agapakis and Tolaas (2012, p. 569) point out that "humans have a remarkably low opinion of the nose, neglecting to cultivate and educate the sense of smell while zealously deodorizing the world." They furthermore note the ubiquity of smells, stating that "we take an average of 24,000 breaths per day, each inhalation swirling countless molecules over our olfactory receptors to give us a smelly glimpse of the chemical world" (p. 569).

It is clear that smell is evolutionary important, affects our daily behaviors in many different ways, and constantly surrounds us. This suggests that the infrequency with which odor concepts are verbalized probably does not exclusively stem from the inferiority of smell as a perceptual process. One possible explanation is then that speakers do not often consciously notice smells. Smeets and Dijksterhuis (2014, p. 7) talk of a "natural inclination" for most people "to pay more attention to visual than olfactory attributes of the environment." Some experimental evidence suggests that participants frequently do not notice changes in odors (Mahmut & Stevenson, 2015; see also discussion in Lorig, 1999). Since humans can only verbalize what they are aware of, this may explain the lack of smell words.[3]

Another option is that speakers notice odors just as much as everything else, but they choose to not talk about them as much, perhaps because of "cultural preoccupation" (Levinson & Majid, 2014, p. 421) such as a cultural disregard of smell, especially compared to sight.[4] Many researchers have noted that vision is culturally dominant in the modern West, with cultural historians and anthropologists thinking of the West as having a vision-centric cultural complex (Classen, 1993, 1997). The converse of this is that smell, together with taste, is often disregarded and treated as inferior. Throughout history, smell has been described as primitive, unimportant, and animalistic by scientists, philosophers, and laypeople (see Classen, Howes, & Synott, 1994; Le Guérer, 2002; Staniewski, 2017; Synnott, 1991). Agapakis and Tolaas (2012), for example, mention that smell has "been intellectually neglected by ancient philosophers and modern art critics alike" (p. 569).

In line with the idea that cultural preferences play a strong role, languages spoken in cultures that emphasize smell more than English-speaking cultures tend to verbalize smell experiences more often or have larger smell vocabularies (Howes, 2002; Lee, 2015; Majid & Burenhult, 2014; San Roque et al., 2015; Wnuk & Majid, 2014). In several cultures from the Andaman Islands, such as the Ongee (Classen, 1993), smell plays an important role in the cosmology and spiritual belief system: aromas are vital energies, death is conceptualized as the loss of one's personal smell, and the year is conceptualized as a cycle of scents (Classen et al., 1994, pp. 95–96). Consistent with the relative importance of smell in their culture, smell occurs in many expressions that Ongee speakers use, such as a conventionalized greeting that translates to 'How is your nose?'. Floyd, San Roque, and Majid (2018) describe the fact that the Barbacoan language Cha'palaa (spoken in Ecuador) has a grammaticized smell classifier *–dyu*. For instance, the smell terms *pindyu* 'sweet smell' or *wishdyu* 'smell of urine' are formed by the roots *pin–* and *wish–*, which crucially cannot stand alone but have to be used with the smell classifier. Classifier systems in the world's languages tend to be for perceptual characteristics such as shape and type of material. So far, Cha'palaa is the only language described to have a grammaticized smell classifier, and, perhaps not surprisingly, smell seems to have special importance for the speakers of this language, who also frequently reference smell concepts in discourse (Floyd et al., 2018). Thus, it appears that cultures in which smell is regarded as more important also have special means to talk about smells.[5]

Thus, when San Roque et al. (2015, p. 50) say that "it could be that there are simply more occasions to talk about visual objects than objects apprehended through the other senses," one has to keep in mind that "more occasions" may arise because of perceptual factors (e.g., vision may be more consciously accessible as a perceptual modality) or cultural factors (e.g., vision may be regarded as more important by speakers). Communicative need not only arises from the way humans perceive the environment directly, but also from how much they are attuned to certain aspects of the environment as a result of cultural belief systems. Currently, it is not known which of the different factors ultimately drives communicative need. However, because there is independent evidence for (a relative) lack of conscious access to smell, as well as independent evidence for the cultural disregard of smell in the West (e.g., Classen et al., 1994), it appears plausible that both perceptual and cultural factors create a diminished communicative need to talk about smell.

Finally, it must be noted that communicative need explanations are suitable for more than clarifying the differential ineffability of the senses; they also easily account for the ineffability of subjective experience, the ineffability of fine perceptual detail, and the ineffability of multisensoriality. Because language is for social coordination and achieving social and physical goals, the details of subjective experience are often communicatively irrelevant. Moreover, most of the time, it is appropriate for speakers to gloss over perceptual detail when talking to others. For example, different shades of the color 'red' can most often be simply labelled *red* without losing our communicative goals. Finally, there are many occasions where the intense multisensoriality of experience is simply irrelevant and we are better served by singling out those aspects of multisensory experience that are relevant to a given task.

4.4.4. Evaluating explanations of ineffability

The present discussion should have made it clear that it is currently unknown what causes the differential ineffability of the senses. This should make us wary of definite statements such as "the reason for the poverty of the lexicon of olfaction is a neurophysiological one" (Holz, 2007, p. 189). There is evidence for the idea that olfaction is neurologically impoverished, and there is separate evidence for the idea that olfaction is linguistically impoverished (see also Chapter 12). However, we cannot make a causal claim based on this mere correspondence of two potentially independent facts. This is because there are multiple possible explanations, as reviewed by Levinson and Majid (2014) and as discussed here. It is at least plausible that phenomena such as the ineffability of smell have joint environmental, neurophysiological, cognitive, linguistic, and communicative causes. In fact, it is at present unreasonable to assume that there is any one cause; a pluralistic explanatory approach may be better suited (cf. Markman, 2008; Mitchell, 2004; Mitchell & Dietrich, 2006), at least as long as there is no definite evidence that lets us decide between the different causal mechanisms.

4.5. Shifting semiotic strategies

Coupling the discussion of the semiotic toolkit (Chapter 3) with the notion of ineffability allows us to see that ineffability is always relative with respect to a particular linguistic strategy. Some semiotic tools are better for some perceptual jobs than others.[6] What makes some semiotic tools more suitable in some situations? Of course, there are constraints that come from the modality of expression, such as limitations on what can be expressed with iconicity in spoken versus signed languages (e.g., Dingemanse, 2013, p. 157; Perlman et al., 2018). Here, however, I want to focus on the role of communicative task demands.

Lu and Goldin-Meadow (2017) show how changing task demands lead to shifts in semiotic strategies (see also Lu & Goldin-Meadow, 2018). These authors asked native English speakers and native ASL signers to describe pairs of objects that differed in their degree of perceptual overlap. More perceptual detail needs to be communicated when the task requires differentiating between two highly similar objects (such as two vases that differ subtly in shape), compared to dissimilar objects (such as a pot and a bowl). Lu and Goldin-Meadow (2017) found that when the task required detailed descriptions, both ASL signers and English speakers relied more strongly on iconic means of expression, and they used less abstract words. That is, iconicity replaced abstract encoding strategies as a function of task demands. This is a clear demonstration of how a particular semiotic strategy is preferred when fine perceptual detail is required.

The results by Lu and Goldin-Meadow (2017) provide a nice demonstration of how the different semiotic strategies have to be considered as a whole: When one tool from the toolkit is not available or insufficient, another tool is chosen. Such trading relations were also seen in a communication game conducted by Roberts, Lewandowski, and Galantucci (2015), who showed that participants were less likely to innovate an abstract encoding strategy if a referent could be communicated iconically. Similarly, it has been observed that the use of abstract color labels in cosmetics publications decreased with the advent of high-fidelity photographing and printing techniques (see Wyler, 2007, p. 123).

Such trade-offs are also discussed in the literature on metaphors. Strik Lievers (2016, p. 52) suggests that in the novel *The Perfume*, in which smell descriptions play a crucial role, the author Patrick Süskind has to resort to metaphor in order to make up for the lack of smell vocabulary. Many researchers think that metaphors are a tool that can be used to make up for lack of vocabulary in a particular perceptual domain (see also Abraham, 1987, p. 156; Engstrom, 1946, p. 10; Holz, 2007; Ronga, 2016; Ronga et al., 2012; Strik Lievers, 2015, 2016; Suárez-Toste, 2013; Ullmann, 1959, p. 283; Velasco-Sacristán & Fuertes-Olivera, 2006). This is in line with Fainsilber and Ortony's (1987) "inexpressibility hypothesis," which states that "metaphors may enable the communication of that which cannot be readily expressed using literal language" (p. 240). The inexpressibility hypothesis can be reconceptualized as a trading relation between semiotic strategies: If there are fewer abstract words for a particular perceptual quality, such as smell, then another strategy is needed, such as borrowing terminology from the other senses.

The perspective of task demands also allows us to reconceptualize expert language, such as the language of sommeliers or the language of music critics. It has been noted that expert descriptions of sensory perceptions may show an increased frequency of source-based language (e.g., Croijmans & Majid, 2016) and metaphor (Barten, 1998; Caballero, 2007; Pérez-Sobrino & Julich, 2014; Porcello, 2004; Suárez Toste, 2007). This can now be seen as an adaptation of expert language to the frequency with which particular task demands arise—namely, the demand of communicating fine perceptual information in detail.

4.6. Conclusion

This chapter discussed a fundamental problem of sensory linguistics, which is the question as to whether certain aspects of sensory experience are ineffable. The discussion focused on the notion(s) of ineffability discussed by Levinson and Majid (2014), essentially providing an update to their seminal paper and reconceptualizing some of the issues they raise.

The amendments to their discussion were as follows: It is important to avoid lumping different kinds of ineffabilities together. Each ineffability highlights a different way through which the sensory world is compressed in language use. It is furthermore important to distinguish the differential ineffability of the senses from other forms of ineffability. Whereas the differential ineffability of the senses is best characterized as a weak form ineffability (some languages have more words for certain senses than others), the ineffability of subjective experience, the ineffability of fine perceptual detail, and the ineffability of the multisensoriality are most likely strong ineffabilities. That is, these three ineffabilities describe aspects of the sensory world that cannot be expressed in any language. Altogether, this discussion detailed the limitations of language, rendering more precise Moore and Carling's (1988, p. 4) statement that "much of the time we expect too much of language."

This chapter discussed how the referential range of sensory words is simultaneously both narrow and broad. It is narrow when considering the particular type of perceptual sensation that is expressed (i.e., only a single perceptual characteristic is highlighted). It is broad when considering that within a given perceptual quality (such as color), most fine-grained perceptual distinctions that are perceivable are not encoded and perhaps not even codable.

At present, we have reason to believe that the differential ineffability of the senses has multicausal origins. For example, vision may be perceptually dominant, neurophysiologically dominant, and culturally dominant. This compels us to take a full suite of possible explanations into account when talking about ineffability, and further work is necessary to lend support to particular explanatory accounts.

Finally, it was argued that whether something is ineffable or not depends on what semiotic strategy is considered, and the different semiotic strategies are weighted differently depending on task demands. The insufficiency of one semiotic strategy can be attenuated by employing a different strategy from the semiotic toolkit. If this happens habitually, expert language develops.

Crucially, however, whatever the semiotic strategy, whatever the perceptual quality, there will always be things that are truly ineffable. Fine perceptual details and the genuinely subjective and multisensory aspects of experience are simply impossible to encode via language.

[1] Interestingly, Levinson and Majid (2014) leave out touch in these descriptions, perhaps because it is difficult to classify touch vocabulary due to its high multisensoriality (e.g., Lynott & Connell, 2009).

[2] In fact, common sensibles pose a problem for all of sensory linguistics, including research on iconicity. For example, the fact that shape characteristics are common sensibles means that it is impossible to know what sensory modality explains the *kiki/bouba* phenomenon discussed in Chapter 3. Visually presented shapes are reliably associated with *kiki* and *bouba*; the same effect is also obtained with shapes that are explored haptically, via touch alone (Fryer, Freeman, & Pring, 2014; see also Fontana, 2013). Making matters worse, humans also have some capacity to determine shapes from sound alone (e.g., Kunkler-Peck & Turvey, 2000; Thoret, Aramaki, Kronland-Martinet, Velay, & Ystad, 2014). This opens up the possibility that when participants match *kiki* with spiky shapes, they might imagine the sound of an angular object when it falls to the ground.

[3] In fact, it is possible that speakers do not notice particular features of their environment *because* they do not have words for those features, or, conversely, that they pay more attention to those aspects of the world that are already verbalized. Thierry and colleagues (Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009) compared the performance of English and Greek speakers in a color detection task. The Greek language differentiates between a darker shade of blue called *ble* and a lighter shade of blue called *ghalazio*, both of which are labelled *blue* in English. Using a color detection task while simultaneously recording event-related brain potentials, these researchers found that Greek speakers exhibited a visual mismatch negativity (vMMN) for the different shades of blue that was absent in English speakers. Such a mismatch negativity is generally thought to indicate preattentive, automatic processing. This shows that Greek speakers subconsciously processed blue colors in line with language-specific conventions. Such results and other results in the emerging field of linguistic relativity research (e.g., Athanasopoulos & Bylund, 2013; Bylund, Athanasopoulos, & Oosendorp, 2013) open up the possibility that *because* there are more words for certain sensory experiences, speakers may be more attuned to paying attention to them.

[4] Speaking of smell, Lorig (1999) proposes that "our limited language for odors may be a cause for our disregard of this sense rather than an effect of getting our noses off the ground" (p. 392). There is some plausibility to this proposal: Given the importance of language in our behavioral ecology, it would seem that speakers may come to be dissatisfied and even disregard smell *because* of the fact that it is difficult to talk about.

[5] Interestingly, Lee (2015) reports that even though the Amis have a relatively large smell vocabulary, they still exhibit the "tip-of-the-nose phenomenon" of finding it difficult or impossible to name certain smells (p. 344). This would point to the possibility that even when speaking a language with a large smell vocabulary, encoding smells is difficult, perhaps because of cognitive reasons (but see Majid & Burenhult, 2014).

[6] It should be emphasized that at present, no systematic quantitative comparisons of the different semiotic strategies involved in perceptual language has been done for controlled stimuli in multiple modalities, with the notable exception of work such as Dubois (2000, 2007). In order to truly know what semiotic encoding strategies are preferred or dispreferred, one needs to control the sensory stimuli experienced by participants.

Chapter 5. The Embodied Lexicon Hypothesis

5.1. Introduction

As the meaning of sensory words is of pivotal interest to sensory linguistics, it is important to talk about different theories of meaning. In particular, this chapter reviews the evidence for the idea that words trigger perceptual simulations, which serves as a motivation to look for correspondences between language use and the behavior of perceptual systems. I will start by talking about modal versus amodal lexical representations and how these are connected to the notions of embodiment, mental imagery, and mental simulation (Chapter 5.2). The evidence for language-induced perceptual simulation leads me to propose the Embodied Lexicon Hypothesis (Chapter 5.3), which I will compare and contrast with other theoretical frameworks in Chapter 5.4. Finally, Chapter 5.5 discusses the fact that not only perceptual representations, but emotional representations as well, are a core part of word meaning. The theoretical framework outlined here will serve as a guiding principle for the data analyses presented in Part II of this book.

5.2. Embodiment, mental imagery, and perceptual simulation

Meaning is one of the most elusive aspects of language, having incited debates for many centuries. The question of meaning turns out to be one of the most fundamental questions in cognitive science, philosophy, and the language sciences. The question of meaning is deeply connected to issues surrounding the nature of our conceptual systems. This chapter focuses on experimental evidence from cognitive science that any linguistic theory of meaning needs to address. The evidence reviewed here will lead me to propose what I call the Embodied Lexicon Hypothesis.

Within the study of meaning, one can distinguish between two major strands: the modal view and the amodal view. The former proposes that concepts in the mind are constituted by modal mental representations—that is, representations that are directly related to the underlying sensorimotor processes involved in perception and action. For instance, the concept of a blender is seen as being constituted, at least in part, by visual representations of a blender (akin to mental images), but perhaps also by auditory representations, such as memory traces of the sounds a blender produces.

The modal view is in stark contrast to the amodal or symbolic view of meaning. According to this view, a word's meaning is an abstract symbol in the mind, akin to the word itself. This suggests the notion of a dictionary filled with unique identifiers. A word such as *bachelor*, then, provides access to the dictionary entry for the corresponding concept. The dictionary entry itself may be composed of such relatively abstract features as [+MALE], [+YOUNG], and [-MARRIED]. Amodal accounts form a natural union with modular approaches to the mind (Fodor, 1983), which see different cognitive faculties as informationally encapsulated. In line with this, amodal accounts view a conceptual representation as separate from processes of perception and action. When amodal accounts allow for connections between abstract symbols and sensorimotor processes, they do so on the assumption that these connections are much less direct (Mahon & Caramazza, 2008) than posited by modal accounts. They do not regard sensorimotor processes as being at the core of semantic representation.

The modal view of meaning will be endorsed in this book and supported by new evidence. The modal view forms a natural marriage with a particular theme discussed in the cognitive sciences: embodiment or embodied cognition (e.g., Anderson, 2003; Barsalou, 1999, 2008; Gibbs, 2005; Glenberg, 1997; Wilson, 2002). Gallese and Lakoff (2005, p. 456) characterize embodiment by stating that cognition and language are "structured by our constant encounter and interaction with the world via our bodies and brains," which includes interaction with the world as it is mediated through the senses. The following definition of embodiment is provided by Evans (2007):

> The human mind and conceptual organisation are a function of the way in which our species-specific bodies interact with the environment we inhabit. In other words, the nature of concepts and the way they are structured and organised is constrained by the nature of our embodied experience. (p. 78)

Embodied approaches to cognition break down the barrier between what is sometimes called "low-level" perception and "high-level" cognition. Willems and Francken (2012, p. 1) say that "embodied cognition stresses that perception and action are directly relevant for our thinking, and that it is a mistake to regard them as separate."

Embodiment, however, is not as much a theory as it is a framework. Different researchers take embodiment to mean different things (see Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Wilson, 2002; Wilson & Golonka, 2013). Embodiment is dealt with in numerous subfields of cognitive science and the language sciences, many of which emphasize different aspects of the overarching framework. In experimental paradigms, diverse embodied effects have been found, often conceptually independent from one another. Furthermore, the embodied cognition framework bleeds into what researchers have variously called grounded cognition, situated cognition, distributed cognition, extended cognition, and dynamical systems approaches to cognition (see Spivey, 2007;

Wilson & Golonka, 2013). These other terms are sometimes used in a way that is synonymous with embodiment, sometimes they are used to emphasize different aspects of language–cognition–brain–body–environment interactions.

The aspect of embodiment that is particularly relevant in the context of this present book are language-external influences on language. Nobody denies that language connects to perceptual processes *in some fashion or other*. However, embodied theories posit deeper interconnections between language and perception. Different theories commit to embodiment in different degrees, forming a continuum from strongly embodied to fully disembodied theories, or from theories that assume fully modal representation to fully amodal representations. Following Meteyard et al. (2012), we may distinguish between a "strong embodiment" view and a "weak embodiment" view. The strong view states that semantic representations are completely dependent on sensorimotor systems. The weak view states that semantic representations are influenced by sensorimotor systems, but they are not wholly constituted by them. According to the weak view, "sensory and motor information does not exhaust semantic content" (Meteyard et al., 2012, p. 799).

Within modal accounts, the specific mechanism that is used to explain the access and retrieval of meaning is claimed to be mental simulation—the idea that language users mentally simulate what a piece of language is about, which involves partially reinstantiating the same neural activation patterns that are involved in actual perception and action (Barsalou, 1999; Bergen, 2012; Fischer & Zwaan, 2008; Zwaan, 2009). Language-induced mental simulation has sometimes been called the "indexical hypothesis" of language understanding (Glenberg & Robertson, 1999, 2000), but it has also been talked about in terms of "perceptual symbol systems" (Barsalou, 1999), or as being part of "grounded cognition"

(Barsalou, 2008). Bergen (2012) provides a particularly vivid introduction to mental simulation accounts of language processing.

Mental simulation is often compared to a re-enactment of perceptual experiences. The proposal is that when language users process concrete language, they mentally activate specific sensory content—relating to vision, touch, hearing, taste, and smell—using the same brain areas that are also responsible for language-independent perception and action (González et al., 2006; Hauk, Johnsrude, & Pulvermüller, 2004; Pulvermüller, 2005). For example, understanding property words such as *loud* and *tart* would be seen as involving the simulation of loudness and tartness experiences (Pecher, Zeelenberg, & Barsalou, 2003). Similarly, a perceptual simulation account would claim that understanding expressions such as *stabbing pain* or *drilling pain* involves the mental simulation of pain experiences (Semino, 2010).

To understand what precisely is meant by mental simulation, it is useful to contrast it with mental imagery (Connell & Lynott, 2016; Bergen, 2012), which is often understood as the deliberate and conscious activation of perceptual content, as when one actively traces a path throughout one's mental map of a city. The heydays of mental imagery research were the 1970s and 1980s, when a lot of the crucial early evidence for the importance of mental imagery was collected. Researchers have found, for example, that the purely mental rotation of 3D blocks has similar characteristics to physically rotating the same blocks (Shephard & Metzler, 1971), and mentally scanning a map with the mind's eye is similar to visually scanning with the blanket eye (Kosslyn, Ball, & Reiser, 1978). This has led to the hypothesis that mental imagery uses the same cognitive and neural resources as actual perception.

Mental simulation is the less deliberate, less conscious, and less vivid version of mental imagery (cf. Connell & Lynott, 2016). When I ask someone to imagine a jar of kimchi with their mind's eye, then this is an instructed process that qualifies as imagery. On the other hand, mentioning the word *kimchi* in passing during a conversation without instructing someone to perform imagery may still trigger a simulation, a less vivid and less consciously accessible form of imagery. As understood here, mental simulation rests on our ability to perform mental imagery.

The evidence presented in this book is framed in terms of simulation, rather than imagery. That is, I will focus on results that support the idea that sensorimotor processes play a role in undirected contexts where a language user may not be consciously performing imagery.

5.3. The evidence for perceptual simulation

There is a wealth of experimental evidence for sensorimotor processes playing a role in language understanding. In the following section, I will review a lot of the experimental and neuropsychological evidence for language-induced perceptual simulation. I should state from the outset that to motivate the Embodied Lexicon Hypothesis in Chapter 5.4 the precise details as to when and how perceptual simulation is involved are not as important as the fact that it is involved. Our focus will be on linguistic patterns, not on different theories of processing. Thus, when discussing the neuropsychological literature in particular, it is not necessary to go into the detailed methodology of each study. For present purposes, it suffices to say that there is a connection between language and perception.

On the motor side of sensorimotor processes, studies have shown that reading or listening to words such as *to kick* or *to push* activates action representations in the brain that are related to what is actually involved in perceiving or performing these actions (Glenberg & Kaschak, 2002; Hauk, Johnsrude, & Pulvermüller, 2004; Pulvermüller, 2005; but see Papesh, 2015 and Tomasino, Fink, Sparing, Dafokatis, & Weiss, 2008). One strand of evidence comes from neuroimaging: The primary motor cortex of the brain is organized somatotopically, which means that there are is a leg-related area, a foot-related area, a mouth-related area, and so on. Hauk and colleagues (2004) conducted an fMRI study which showed that reading words such as *kick* increases blood flow in areas that are commonly associated with leg-related actions, while verbs such as *to push* or *to hit* lead to relatively more blood flow in hand-related areas.

The sensory side of sensorimotor processes is more relevant for the present book. Here, vision has by and large received the most attention. In a classic experiment, Stanfield and Zwaan (2001) showed that participants reading a sentence such as *John pounded the nail into the wall* more quickly responded to a visual image of a nail that is horizontal, because pounding a nail into a wall implies a perpendicular orientation to a vertical wall. On the other hand, participants reading the sentence *John pounded the nail into the floor* responded more quickly to a visual image of a vertical nail perpendicular to a horizontal surface. This sentence–picture match effect was taken to suggest that reading a sentence activates the visual representation of a nail, which is specific enough to encode the spatial orientation implied by the sentence (for replication and extension studies, see Pecher, van Dantzig, Zwaan, & Zeelenberg, 2009, and Zwaan & Pecher, 2012).

Similar match effects have also been obtained for color: Mannaert, Dijkstra, and Zwaan (2017, Experiment 1) asked participants to read sentences such as *The driving instructor told Bob to stop at the traffic lights* and *The driving instructor told Bob to go at the traffic lights*. Participants were faster to verify the visual image of a red traffic light for the "stop" sentence and the visual image of a green traffic light for the "go" sentence, suggesting that they had formed a perceptual representation of color during reading. More evidence for the mental simulation of color is presented by Connell (2007) and Connell and Lynott (2009).

Numerous other aspects of visual experience have been shown to matter in mental simulation. Besides color and spatial orientation, similar experiments have found evidence for the perceptual simulation of visual shape (Zwaan & Pecher, 2012; Zwaan, Stanfield, & Yaxley, 2002), direction of motion (Kaschak et al., 2005; Meteyard, Bahrami, & Vigliocco, 2007; Zwaan et al., 2004), distance (Vukovic & Williams, 2014; Winter & Bergen, 2012), size (de Koning, Wassenburg, Bos, & van der Schoot, 2016), and clear versus foggy or murky visibility conditions (Yaxley & Zwaan, 2007).

What about the other senses? What evidence is there for language-induced mental simulation for the non-visual modalities? For sound, Winter and Bergen (2012) showed that participants react faster to a relatively loud sound of a bleating sheep after reading *The sheep walks up to you and bleats* compared to a quiet sound, which participants respond faster to after reading *The sheep wanders to the other side of the hill from you and bleats*. Kaschak, Zwaan, Aveyard, and Yaxley (2006) further demonstrated that language comprehenders mentally simulate the direction of auditory motion. Kiefer, Sim, Herrnberger, Grothe, and Hoenig (2008) found processing visually presented words that strongly relate to sound impressions, such as *telephone*, involved activity in the left posterior gyrus (pSTG) and middle temporal gyrus (pMTG). When participants in the same study attentively listened to recordings of sound events, such as animal sounds, the same pSTG/pMTG cluster was involved. This suggests that perceptual and conceptual processes converge on the same neuroanatomical regions. Moreover, comparison to a behavioral rating study showed that involvement of pSTG/pMTG increased linearly with the perceived relevance of acoustic properties to the corresponding words (i.e., words that more strongly relate to sound more strongly engaged sound-related brain areas).

For touch, a neuroimaging study by Lacey and colleagues found that reading metaphors involving texture-related touch words, such as *She had a rough day*, led to increased blood flow in texture-sensitive regions of somatosensory cortex (the parietal operculum) in comparison to reading similar literal expressions such as *She had a bad day* (Lacey, Stilla, & Sathian, 2012).

Given that pain can be considered a sense (nociception), evidence for perceptual simulation also comes from studies of pain words: Reading pain-related words such as *drilling*, *excruciating*, and *grueling* leads to increased blood flow in the brain's pain matrix in comparison to other negative words such as *disgusting* or *scary* (Richter, Eck, Straube, Miltner, & Weiss, 2010). Osaka, Osaka, Morishita, Kondo, and Fukuyama (2004) furthermore investigated Japanese ideophones expressing pain, such as *kiri-kiri* to depict a stabbing pain, *zuki-zuki* for a throbbing pain, or *chiku-chiku* for an intermittent pain. These researchers found that listening to such words compared to nonsense syllables leads to increased blood flow in the anterior cingulate cortex, an area known to be involved in registering actual pain sensations (see also discussion in Semino, 2010). Vukovic, Fardo, and Shtyrov (under review) found that reading literal descriptions of pain leads to increased sensitivity toward subsequent pain stimulation, especially for patients with chronic pain.

There is also extensive evidence for gustatory simulation. Barrós-Loscertales and colleagues found increased blood flow in primary and secondary gustatory cortices when Spaniards read taste-related words such as *cerveza* 'beer' and *chorizo* 'spicy sausage' in contrast to control words with little gustatory association, such as *casco* 'helmet' and *piel* 'skin' (Barrós-Loscertales et al., 2011). Citron and Goldberg (2014) found increased blood flow in primary and secondary gustatory areas when Germans read metaphorical statements involving taste words (e.g., *Sie bekam ein süsses Kompliment* 'She received a sweet compliment') as opposed to corresponding literal statements (e.g., *Sie bekam ein nettes Kompliment* 'She received a nice compliment').

More evidence for the involvement of gustatory brain areas in processing taste-related language comes from so-called property verification tasks, where participants are asked to answer questions such as "Can cranberries be tart?" Goldberg, Perfetti and Schneider (2006a) asked participants to verify when an object possessed a visual property (e.g., *green*), an auditory property (e.g., *loud*), a tactile property (e.g., *soft*), or a taste property (e.g., *sweet*). They showed that verifying gustatory properties led to increased blood flow in the left orbitofrontal cortex, an area previously shown to be involved in flavor processing. Similarly, retrieving tactile knowledge was associated with increased blood flow in somatosensory, motor, and premotor areas (see also Goldberg, Perfetti, & Schneider, 2006b).

Evidence for the involvement of smell-related brain areas in processing smell-related words comes from González and colleagues (2006). These researchers found increased blood flow in the piriform cortex and the amygdala when Spaniards read odor-related words such as *aguarrás* 'turpentine' and *orina* 'urine' as opposed to control words such as *sierra* 'saw' and *mago* 'wizard' (González et al., 2006). Importantly, the same brain areas were previously found to be involved in smelling odors (Zatorre, Jones-Gotman, Evans, & Meyer, 1992). Pomp and colleagues report increased blood flow in secondary olfactory areas (orbitofrontal cortex) when participants read metaphorical and literal sentences involving smell words; however, they did not find increased blood flow in primary olfactory cortex (piriform cortex; Pomp et al., 2018).

Some dissenting evidence for olfactory simulation comes from Speed and Majid (2018). This study involved two experiments: one looking at auditory simulation, another one looking at olfactory simulation. Findings showed that sounds were more likely to be remembered when paired with a congruent word (e.g., the sound of bees buzzing paired with the word *bee*). Moreover, when remembering sound words, recall was slower when paired with incongruent sounds (e.g., the word *typhoon* paired with a bee sound). These interactions between language and actual sound stimuli provide further evidence for the idea that language about sounds engages auditory simulations. On the other hand, Speed and Majid (2018) did not find similar effects when odor language was paired with actual odors (i.e., memory of words or odors was not impaired when the two mismatched). However, odor-related language did affect immediately following judgments of odor intensity and pleasantness, suggesting that there are perceptual representations at some level of cognitive analysis, albeit perhaps coarser ones than in the case of auditory representations for sound-related words.

The evidence for perceptual simulation presented so far involved neuroimaging and behavioral experiments. On the behavioral side, the experiments often relied on match or interference effects: The logic is that if language engages perceptual processes, language should be able to interact with behavior in a task that is purely perceptual in nature, either by facilitating task performance or by interfering with it (Bergen, 2007). However, there is a whole other class of experiments that provide evidence for the engagement of sensorimotor processes in language understanding, although in a less direct fashion. These experiments support embodied semantic representations by pointing to analogous behavior in perceptual and linguistic tasks. These experiments rely on what I call the "correspondence argument": If sensory language engages perceptual representations, then processing said language should mirror perceptual effects involving the described senses.

An example of the correspondence argument is given by Connell and Lynott (2012), who found a tactile disadvantage in the conceptual processing of sensory words such as *chilly* and *stinging*. This psycholinguistic finding corresponds to independent perceptual evidence where a tactile disadvantage has previously been established for actual tactile stimulation (see Spence et al., 2001; Turatto et al., 2004; see also Karns & Knight, 2009).

Another example of the close correspondence between relatively high-level phenomena and perceptual processes comes from Pecher, Zeelenberg, and Barsalou (2003). Again, participants were asked to verify whether an object has a certain property—for example, *a blender can be loud* (true) versus *an oven can be baked* (false). Pecher and her colleagues (2003) found that when participants verified a property in one modality, such as auditory (*blender–loud*), they were subsequently slower when performing a judgment in a different modality (*cranberries–tart*). There was no similar interference when performing a judgment in the same modality (*leaves–rustling*). Thus, the trial sequence "*blender–loud → leaves–rustling*" resulted in faster responses than the trial sequence "*blender–loud → cranberries–tart*" (Connell & Lynott, 2011; Louwerse & Connell, 2011; Lynott & Connell, 2009; van Dantzig, Cowell, Zeelenberg, & Pecher, 2011; van Dantzig, Pecher, Zeelenberg, & Barsalou, 2008). Importantly, this so-called "modality switching cost" is not confined to just words; it was previously shown to characterize switching between perceptual modalities in a purely nonlinguistic task (Spence et al., 2001; Turatto et al., 2004). For instance, hearing a beep after seeing a light flash results in slower detection of the light flash compared to seeing two light flashes in a row. Thus, there is a modality switching cost in perception as well as in the linguistic processing of perceptual words, which provides indirect evidence for the idea that language engages modality-specific representations.

5.4. The Embodied Lexicon Hypothesis

The key proposal defended in this book is that the linguistic behavior of sensory words such as *salty* and *silky* can be partially explained by how the senses differ from each other in perception, and by how the senses interact with each other in the brain and behavior. This fundamental idea is nicely summarized in the statement that "properties of sensory experience wend their way through language—permeating that most human manifestation and expression of thought" (Marks, 1978, p. 3).

The Embodied Lexicon Hypothesis is an adaptation of this general view. In its most general form, it says that language mirrors perception. More specifically, it involves the following two sub-hypotheses (compare Marks, 1978, p. 3):

(a) Perceptual asymmetries result in linguistic asymmetries.

(b) Perceptual associations result in linguistic associations.

The only commitments of the Embodied Lexicon Hypothesis are that (a) words activate sensory-motor representations, and (b) those sensory-motor representations partly determine word choice. As a result of this, (c) language comes to reflect sensory-motor processes in its structure, as well as in language use. Thus, the Embodied Lexicon Hypothesis encompasses both processing and linguistic patterns that occur in natural language. However, because there is already a lot of evidence for simulation being involved in processing (as reviewed above), the present book focuses on establishing that linguistic patterns, such as evidenced through analyses of corpora, also follow embodied principles.

The Embodied Lexicon Hypothesis crucially rests on the correspondence argument and thus calls for looking at perceptual evidence in relation to new linguistic evidence. This was already discussed in the context of such findings as the "tactile disadvantage," which was found in both word processing (Connell & Lynott, 2012) and perception (Spence et al., 2001; Turatto et al., 2004). The novelty here is to shift the focus of the correspondence argument to issues of linguistic *structure and use.*[1] The Embodied Lexicon Hypothesis accepts the evidence for the involvement of perceptual simulations in word processing, but it goes one step further by stating that this does not leave linguistic patterns untouched. That is, the ways that sensory words are used should reflect patterns independently observed in perception *precisely because* words engage perceptual processes.

Trivially, every theory of meaning, modal or not, would agree that the things we perceive need to be encoded in the sensory vocabulary in some fashion. For example, English speakers have words for such properties as *sweet* and *red* because our perceptual apparatus is able to perceive these properties. The converse is true as well: We cannot have sensory words for things that we cannot perceive. Language focuses on the already filtered, relatively coarse aspects of the natural world that are the result of sensory perception, and the perceptual acuity of our sensory systems sets outer limits to the levels of detail that can possibly be encoded in language. From this perspective, the Embodied Lexicon Hypothesis is evidently true in a trivial fashion (i.e., language and perception have to be related somehow). However, the proposal goes beyond this, stating that the fit between language and perception is perhaps greater than some theories of language would admit. This book will showcase many distributional patterns that are best explained with recourse to language-external perceptual processes, in line with the Embodied Lexicon Hypothesis.

I should like to make clear that the Embodied Lexicon Hypothesis does not state that language and perception are isomorphic. The mapping between perception and language is far from complete, as was already discussed in Chapter 4 regarding ineffability. In fact, the very idea of ineffability requires that the mapping between language and perception is not perfect. Compared to our subjective, multisensory, and high-resolution experience of the world, language is compressive, digital, and sequential. This means that language and perception can never be fully aligned.

However, the lack of posited isomorphism between language and perception also means that the Embodied Lexicon Hypothesis is, unfortunately, a very weak theory in a Popperian sense.[2] It is difficult to falsify, and many of the findings that will be discussed as supporting the Embodied Lexicon Hypothesis are consistent with some versions of amodal symbol theories. Several authors have criticized how cognitive scientists deal with the evidence that supports embodied approaches to cognition, including those that fall within the embodied camp (Willems & Francken, 2012).

Mahon and Caramazza (2008) argue that activation of sensorimotor systems is downstream in cognitive terms, meaning that the engagement of sensorimotor processes happens *after* symbolic or amodal processing. The idea here is that the concept of a tree may be an amodal symbol that ultimately links up with perception, but only after the fact. Embodied effects in processing are then explained away as being due to indirect links between amodal symbols and perceptual systems. This view renders embodied experimental results epiphenomenal.[3] It is impossible to craft an argument for the necessity of perceptual representations based on linguistic evidence alone. We can, however, make a different (and somewhat weaker) argument, following the discussion in Stanfield and Zwaan (2001) and Zwaan et al. (2002): The amodal symbol systems account can only *postdict* the linguistic patterns discussed in this book, explaining correspondences between perception and language after the fact. The Embodied Lexicon Hypothesis is able to *predict* these correspondences.[4]

The Embodied Lexicon Hypothesis presented here is furthermore "consilient" (Wilson, 1998), in the sense of being able to explain a diverse number of facts. In general, a theory that can explain more distinct classes of facts is to be preferred over one that explains fewer (Thagard, 1978). Amodal approaches leave large bodies of evidence hanging next to each other—namely, facts about perception next to highly related facts about the linguistic patterns of sensory words. It is neither parsimonious nor theoretically elegant to leave two large bodies of empirical evidence about similar (and obviously related) constructs without theoretical integration. Thus, at least in part, the Embodied Lexicon Hypothesis is aimed at synthesis of already existing results.

5.5. Relations to other theories

Several of the patterns discussed in this book under the banner of the Embodied Lexicon Hypothesis are in line with a particular branch of the language sciences that is called "cognitive linguistics" (Croft & Cruse, 2004; Evans & Green, 2006; Langacker, 1987, 2008; Lakoff, 1987; Lakoff & Johnson, 1980). Just like embodiment, cognitive linguistics is not a unified theory; it is a loose framework of interconnected theories and hypotheses that relate to each other through a shared set of assumptions.

Among other things, cognitive linguistics sees meaning as embodied: Semantic structure is argued to be based on conceptual structures (Evans, Bergen, & Zinken, 2007), and conceptual structures are seen as directly relating to perceptual structures and patterned interactions with the physical environment. For example, people talk of *high numbers* or *rising prices* because they have an embodied understanding of quantity in terms of the concrete physical concept of vertical space (Winter et al., 2015a). This understanding is embodied because it is presumed to have arisen through a lifetime of interactions with physical quantities, which literally pile up vertically (Lakoff, 1987, p. 276).

The present book can be seen as being loosely affiliated with the tradition of cognitive linguistics. However, in contrast to many cognitive linguistic studies, the focus here is on large-scale quantitative aspects of lexical structure. Cognitive linguistics, like any subfield within the language sciences, focuses on a certain type of data and a certain style of analysis (such as the analysis of time metaphors, among many other topics). While perceptual language is occasionally discussed within cognitive linguistics (e.g., Caballero, 2007; Caballero & Ibarretxe-Antuñano, 2014; Ibarretxe-Antuñano, 2008; Matlock, 1989; Sweetser, 1990; Yu, 2003), it is not always the primary focus. Thus, the present book can be seen as extending the core idea of embodiment—which is generally seen as part of the cognitive linguistic framework—to new empirical domains, such as the composition of the lexicon or the statistical patterns of word usage. The Embodied Lexicon Hypothesis thus forms a natural marriage with cognitive linguistic approaches to studying language.

The Embodied Lexicon Hypothesis is furthermore related to certain theories within cognitive science. Several researchers who can be seen as part of the embodied tradition have begun to reintroduce a role for amodal, symbolic, or purely linguistic processes into their theories of language comprehension. This is the case with Barsalou and colleagues' (2008) Language and Situated Simulation (LASS) framework, as well as in Louwerse's Symbol Interdependency Theory (Louwerse, 2011; Louwerse & Connell, 2011; see also

Meteyard et al., 2012). These theories assume that when a word is read, associated words are immediately activated. Thus, a specific piece of language may not only be understood in terms of embodied representations, but also in terms of connections to other linguistic representations.

In his Symbol Interdependency Theory, Louwerse (2011) argues that mental simulation results, such as Pecher et al.'s (2003) modality switching effect (discussed above), can partially be explained by linguistic associations rather than embodied associations. For example, the switch from *leaves–rustling* to *blender–loud* may be fast not because both properties are auditory, but because the corresponding words are mentally associated with each other (see evidence in Louwerse & Connell, 2011). Thus, according to Louwerse (2011), the fact that linguistic items are associated with each other influences language understanding above and beyond what comes from embodiment alone.

However, it should be noted that the Symbol Interdependency Theory can only explain embodied cognition results via linguistic associations if those associations represent embodied information in the first place. Thus, embodiment influences processing two ways: first, directly through the activation of sensorimotor content in perceptual simulation, and second, indirectly via feedback from the linguistic system which has already encoded embodied patterns. The Embodied Lexicon Hypothesis focuses on how these linguistic associations arise, and whether they do indeed correspond to perceptual patterns.

In the next section, I will look at how ideas surrounding the Embodied Lexicon Hypothesis and mental simulation relate to another important dimension of meaning: the emotional dimension.

5.6. Emotional meaning

The mental lexicon not only represents perceptual meaning, but also what I will loosely call "emotional meaning," or the "emotionality" of taste words. There are many different cognitive models of emotion, as well as a host of accounts of how emotion (e.g., Wilce, 2003) and, more generally, evaluation (e.g., Hunston, 2010) are performed in language. In the context of this book, I will ignore distinctions between different emotional qualities (such as fear, anger, happiness, etc.) and consider only one aspect of emotion, defined as what is generally called "emotional valence" or merely "valence" in the psycholinguistic literature (Warriner, Kuperman, & Brysbaert, 2013). Valence as used here is understood to be the positive–negative dimension of meaning—that is, whether a word generally expresses something good (*pleasant, tasty, wonderful*) or bad (*disgusting, horrible, vile*). Alternative words for "valence" could be "positivity" or "negativity." Words that are more strongly positive or negative will be talked about as relatively more "valenced," "emotional," or "evaluative" words, compared to relatively more neutral words. I do not neglect the fact that there are other important aspects of emotional meaning, but for present purposes, it suffices to focus on the positive–negative dimension of evaluative and affective language (for further justification, see Chapter 10).

Evidence for the fact that emotional meaning is encoded in lexical representations comes from the fact that a word's emotional valence affects its processing speed in reading and naming tasks (Kousta, Vinson, & Vigliocco, 2009; Kuperman, 2015; Kuperman, Estes, Brysbaert, & Warriner, 2014; Snefjella & Kuperman, 2016). For example, positive words are processed more quickly than negative words, and strongly positive and strongly negative words are overall processed more quickly than neutral words (Kuperman, 2015). Another task suggesting that valence is part of a word's lexical representation is the phenomenon of affective priming, in which a positive word is read more quickly after another positive word, and negative words are read more quickly after negative words (Fazio, Sanbonmatsu, Powell, & Kardes, 1986; de Houwer & Randell, 2004). The evidence from affective priming suggests that emotional meaning is encoded in the lexical representations of words.

There is also evidence suggesting that processing emotion-laden language may involve mental simulation of emotional reactions (for review, see Havas & Matheson, 2013). Havas and colleagues (2007) instructed participants to hold a pen in their mouth, either using just the lips to form a pouting face, or using just the teeth to form a smiling face. Previous research showed that the corresponding facial positions actually make people feel better (smiling condition) or worse (pouting condition; Strack, Martin, & Stepper, 1988; but see Wagenmakers et al., 2016). Havas and colleagues (2017) showed that participants read pleasant sentences such as *You and your lover embrace after a long separation* faster in the teeth condition, when making a smiling face.

More evidence for emotion simulation comes from Foroni and Semin (2009), who used electromyography (EMG) to show that muscles commonly involved in smiling were relatively more activated when reading positive words, andmuscles involved in frowning were more strongly activated for negative words. Subsequently, Havas and colleagues (Havas, Glenberg, Gutowski, Lucarelli, & Davidson, 2010) showed that paralyzing muscles involved in expressing emotions of anger and sadness selectively affected the comprehension of sentences involving these emotions compared to sentences about happy emotions. There is also considerable evidence from neuroimaging that reading emotional words involves increased blood flow in the amygdala (reviewed in Citron, 2012), a brain area also involved in emotion processing. Together, these studies suggest that any account of embodied semantics is incomplete without considering the role of emotion simulation, alongside perceptual simulation.

In addition to these general considerations, the fact that emotions are represented in the mental lexicon and engaged during language use is important for specific results that will be presented in this book under the banner of the Embodied Lexicon Hypothesis. First, I will show that differences in how the senses relate to emotional processes correspond to how sensory words are used evaluatively (Chapter 16). In particular, I will argue that the emotionality of taste and smell words can be explained with recourse to how taste and smell as perceptual processes are tied to emotional processes in the brain and in behavior (see also Winter, 2016). To account for this result, the Embodied Lexicon Hypothesis needs to allow for emotional meaning to be part of a word's lexical representation, alongside perceptual meaning.

In addition, considering the emotional dimension next to the perceptual is important because there may be a trade-off between the two dimensions in the English lexicon. Vigliocco, Meteyard, Andrews, and Kousta (2009) suggest that abstract words such as *freedom*, *courage*, *government*, and *dignity* are more strongly characterized by their emotional meaning than concrete words, which may not need emotional information as a crucial part of their representation because their meaning is already supported by perceptual representations. This view is corroborated by empirical evidence: Kousta, Vigliocco, Vinson, Andrews, and Del Campo (2011) show that abstract words are more emotional than concrete words. Meteyard et al. (2012, p. 800) thus suggest that abstract knowledge may be "grounded in our internal experience," including our emotions, whereas "concrete knowledge would be grounded in our experience with the outside world." This trading relation between perceptual and emotional content has relevance for discussions of so-called "synesthetic metaphors" (Chapters 6–9).

In general, the evidence reviewed in this chapter clearly shows that both perceptual representations and emotional representations are part of word meaning. The Embodied Lexicon Hypothesis recognizes these facts and describes how perceptual meanings relate to each other in language, as well as how perception and emotion are interrelated. Thus, engaging with the Embodied Lexicon Hypothesis calls for a joint consideration of both perceptual meaning and emotional meaning.

---

[1] Of course, processing and structure are not independent of each other, and they interact in many ways (e.g., Christiansen & Chater, 2008; Hawkins, 2004; O'Grady, 2005).

[2] In terms of Meteyard et al. (2012), the findings reported in this book that are used to argue for the Embodied Lexicon Hypothesis are consistent with a strong embodiment view where semantic representations are completely dependent on sensory-motor systems, as well as with a weak embodiment view where semantic representations have partial dependence on those systems. Because the evidence discussed in this book is linguistic in nature, it will be impossible to distinguish between weak and strong embodiment, and distinguishing between these two proposals is not the main goal.

[3] It is important to note that there is already experimental evidence inconsistent with a fully epiphenomenal view of embodied representations, such as effects showing that language processing is influenced by perceptual processes (rather than the reverse, as is the case in the sentence–picture matching task); see, for example, Kaschak et al. (2005) and Pecher et al. (2009). Moreover, there are studies showing that changing the activation level of the motor system via transcranial magnetic stimulation of the brain influences the processing of action-related language, which further suggests a function role (e.g., Willems, Labruna, D'Esposito, Ivry, & Casasanto, 2011).

[4] Moreover, as a research program for sensory linguistics, the Embodied Lexicon Hypothesis invites us to actively look for new perception-language correspondences, which is not something that is done within the amodal tradition.

Chapter 6. Synesthesia and metaphor

6.1. Introduction to synesthesia

A core topic in sensory linguistics is the fact that sensory words can be used flexibly, including in contexts that do not relate to their dominant sensory modality. This happens, for example, when an English speaker says that a particular sound is *rough* or *smooth*. These two adjectives are strongly associated with the tactile modality; hence, their application to the auditory modality is understood by many linguists and literature scholars to be metaphorical—a mapping between two distinct sensory modalities. Related examples include *smooth taste* (touch-to-taste), *rough smell* (touch-to-smell), *sharp sound* (touch-to-sound), *sweet melody* (taste-to-sound), *bright sound* (sight-to-sound), and *loud color* (sound-to-sight). Such expressions have variously been labelled "synaesthetic metaphors" (Ramachandran & Hubbard, 2001; Strik Lievers, 2016), "linguistic synaesthesia" (e.g., Holz, 2007; Whitney, 1952), "verbal synaesthesia" (Strik Lievers, 2015), "literary synesthesia" (e.g., O'Malley, 1957), and "poetic synesthesia" (Shen, 1997, 1998). In literature studies, synesthetic metaphors are often just called "synesthesia" (Engstrom, 1946).

The term synesthesia is a combination of the Greek morphemes *syn* for 'together' and *aisthēsis* for 'sensation.' Most generally, synesthesia refers to any union of the senses. Deroy and Spence (2013, p. 643) criticize the fact that the label synesthesia "acts as something of a placeholder with which to characterize the process (or processes) that underlie surprising reports of associations between two apparently disjoint sensations, categories, or sensory dimensions."

In the clinical literature, synesthesia is generally understood to be a relatively rare neuropsychological phenomenon where a stimulus from one sensory modality, called the inducer, automatically and involuntarily triggers sensations in another sensory modality, called the concurrent (Martino & Marks, 2001; Ramachandran & Hubbard, 2001; Simner, 2006, 2012; Simner et al., 2012). Following Deroy and Spence (2013), this phenomenon will be called "canonical synesthesia" for the remainder of this chapter. An example of synesthesia is colored hearing synesthesia, where a synesthete sees colors in front of their inner eye when hearing certain sounds. Are such phenomena related to synesthetic metaphors, and if so, how?

This chapter will move from more detail on synesthetic metaphors (Chapter 6.2) to a discussion of canonical synesthesia versus crossmodal correspondences (Chapter 6.3). After that, we are in a position to evaluate the connection between synesthetic metaphors and canonical synesthesia (Chapter 6.4), if there is any at all.

6.2. Characterizing synesthetic metaphors

Ullmann (1959) is commonly cited as one of the most seminal studies on the extended meanings of sensory words. In his earlier 1945 paper, *Romanticism and Synaesthesia,* Ullmann analyzed expressions such as *Taste the music of that vision pale* (p. 815, from John Keats' *Isabella*). According to Ullmann, this line involves two "transfers" that cross the senses—namely, from taste to sound (*taste the music*), and from sound to sight (*music of that vision pale*).

The directionality of the transfer is gleaned from the grammatical construction. For example, the adjective–noun pair *sharp sound* is supposed to be a touch-to-sound mapping because the syntactic head of this noun phrase is sound-related and the modifying adjective is touch-related (see e.g., Ronga et al., 2012, p. 145). Because sound is associated with the head of the adjective–noun, this expression is seen as being primarily about sound rather than about sharpness. Thus, semantically, sound is treated as the target of synesthetic transfer, and touch is treated as the source. Research in this field often focuses on adjective–noun pairs and generally adopts the perspective of the adjective modifying the head noun, rather than the related perspective of the head noun narrowing down the sensory meaning of the adjective (cf. Abraham, 1987, p. 161).

Ullmann (1945) uses the term synesthesia as it is frequently used in literature studies, where synesthesia is seen as a particular type of figurative language—a trope that involves "transfers" from one sense to another. Below, I list a range of definitions from literature on these expressions:

> "Synesthetic metaphors...are expressions in which words or phrases describing experiences proper to one sense modality transfer their meanings to another modality."
> (Marks, 1982a, p. 177)

> "Synesthetic metaphor employs the language of one sensory or perceptual domain to transfer meaning to another domain."
> (Marks, 1982b, p. 15)

> "A synaesthesia is a metaphorical expression in which the source and target domains represent concepts belonging to two different modalities or senses."
> (Shen & Cohen, 1998, p. 124)

> "A synaesthetic metaphor is a systematic relationship between elements from two distinct sensory modalities."
> (Shen & Gil, 2007, p. 51)

> "In synaesthetic metaphors, words that pertain to one sensory modality (e.g., vision) are extended to express another sensory modality (e.g., audition)."
>
> (Cacciari, 2008, p. 427)

> "...a perceptual experience related to one sense is described through lexical means typically associated with a different sense..."
>
> (Strik Lievers, 2015, pp. 69-70)

> "Linguistic synaesthesia is a particular form of metaphor, as it extends the meaning of an utterance from one sensory modality to another, through analogy."
>
> (Ronga et al., 2012, p. 139)

The reason why I list so many definitions is to highlight that there are certain assumptions inherent to most research on synesthetic metaphors. In particular, definitions tend to evoke some form of transfer or mapping between two sensory modalities, as when Yu (2003) says that a synesthetic metaphor is a "metaphor that maps across various sensory domains" (p. 20). In addition, the definitions frequently mention that there are two "distinct" sensory modalities. Other definitions highlight more strongly that there needs to be some incompatibility or perceived conflict between the two perceptual sensations that are combined with each other:

> "...synaesthesia is the syntactic relation between elements semantically incompatible, denoting sensations from different sensorial spheres."
>
> (Erzsébet, 1974, p. 25)

> "...in synaesthetic expressions, syntactic links between sensory lexemes create connections that generate conflict at the conceptual level."
> (Strik Lievers, 2016, p. 45)

A useful distinction was introduced by Werning, Fleischhauer, and Beseoglu (2006) and Petersen, Fleischhauer, Beseoglu, and Bücker (2008), who differentiate between "weak" and "strong" synesthetic metaphors (see also Abraham, 1987, p. 179; Engstrom, 1946, p. 11). According to them, a weak synesthetic metaphor has a perceptual source and a relatively more abstract target (e.g., *cold anger*), whereas a strong synesthetic metaphor has both a perceptual source and a perceptual target (such as *cold smell*). Here, I will focus first and foremost on strong synesthetic metaphors. A defining feature of these metaphors is that both the source and the target are perceptual (see Shen, 2008, p. 302).

It is important to stress that the term "synesthetic metaphor" has two components, "synesthetic" and "metaphor." Each one of these components reflects certain assumptions about the underlying phenomenon. First, calling expressions such as *rough sound* and *sweet fragrance* "synesthetic" suggests that there is some connection to what psychologists and neuroscientists call synesthesia. Second, calling these expressions "metaphorical" reflects the assumption that there is a mapping between two distinct modalities.

This chapter questions the assumption that synesthetic metaphors are synesthetic. The next chapter questions the assumption that synesthetic metaphors are metaphorical (see also Winter, in press). I will conclude that synesthetic metaphors are neither synesthetic nor metaphorical.

6.3. The importance of terminology

From its outset, research on synesthesia was fraught with terminological confusion (see Deroy & Spence, 2013). O'Malley (1957) details some of the early work on synesthesia as a neuropsychological phenomenon, showing how the phenomenon has attracted much speculation and confusion in poetry, arts, and in psychology. Often, different researchers take the term "synesthesia" to mean different things. This point was also made by Martino and Marks (2001):

> Over the two centuries since strong synesthesia was first identified in the scientific literature, several heterogeneous phenomena have been labeled as synesthetic. These phenomena range from strong experiences..., on the one hand, to weaker crossmodal literary expressions, on the other. We believe it is a mistake to label all of these phenomena simply as synesthesia because the underlying mechanisms cannot be identical... (p. 62)

This quote also highlights why it may be important to worry about terminological confusion in this domain. Namely, shared terminology suggests shared mechanism; it suggests that two phenomena are more similar to each other than they are different from each other, which in turn may bias research in the direction of seeking similarities rather than differences (see Deroy & Spence, 2013).

Deroy and Spence (2013) highlight that not all sensory interactions qualify as synesthesia. In particular, they highlight the importance of distinguishing between canonical synesthesia on the one hand and crossmodal correspondences on the other. In contrast to canonical synesthesia, the term "crossmodal correspondence" (see Spence, 2011) refers to "the tendency for a feature, or attribute, in one sensory modality to be matched (or associated) with a sensory feature, or attribute in another sensory modality" (Spence, 2012, p. 37). Crossmodal correspondences are perceptual associations that are widespread in the general population, may have environmental origins (e.g., repeatedly experiencing certain correlations between perceptual qualities, such as between small size and high pitch), and are not necessarily consciously accessible. An example of a crossmodal correspondence is the fact that people systematically match brighter visual stimuli to higher pitched sounds and darker visual stimuli to more lower pitched sounds (Marks, 1982a), without consciously perceiving sound sensations when seeing color (or vice versa).

Deroy and Spence (2013) highlight the rarity of synesthesia in the general population and the role of consciousness in synesthetic perceptions as two factors that help demarcate synesthesia from crossmodal correspondences. This is reflected in the definition of synesthesia by Grossenbacher and Lovelace (2001), who describe it as "a *conscious* experience of systematically induced sensory attributes that are *not experienced by most people* under comparable conditions" (p. 36, emphasis in original). In contrast, crossmodal correspondences are experienced by everybody, and they are generally not experienced consciously. For some researchers at least, the rarity of synesthesia is almost definitional; that is, because crossmodal correspondences are perceived by most people, they do not qualify as genuine synesthesias.

6.4. Canonical synesthesia and metaphor

Unfortunately, researchers studying sensory language have not always been clear about how they interpret the term "synesthesia" in the context of metaphor. Different researchers take "synesthesia" as a linguistic phenomenon to be more or less related to "synesthesia" as a perceptual phenomenon, sometimes without explicitly stating where they stand (see already the discussion in O'Malley, 1957).

Moreover, researchers studying sensory language have often not been precise about what they mean by "synesthesia" on the perceptual side, often conflating it with crossmodal correspondences.

It is interesting to note that several researchers studying metaphor introduce the term "synesthesia" followed by some form of hedging, or at least some clarification to differentiate it from canonical synesthesia as a neuropsychological phenomenon. Tsur (2008) states that "the term *synaesthesia* suggests the joining of sensations derived from different sensory domains," and then follows this with a call to "distinguish between the joining of sense *impressions* derived from the various sensory domains, and the joining of *terms* derived from the *vocabularies* of the various sensory domains" (p. 283, emphasis in original). He then says that in synesthetic metaphors it is only the "*terms* that are derived from two sensory domains." Engstrom (1946) stresses that the employment of "synesthesia" as a term for a stylistic trope in literature is disconnected from the psychological phenomenon. Ronga (2016), like many others before her, stresses that "linguistic synaesthesia has to be distinguished from *perceptual* synaesthesia" (p. 48, emphasis in original). She furthermore highlights that "the two synaesthetic phenomena are very different" (see also Cazeaux, 2002, pp. 3–4). In fact, all of these statements echo Ullmann (1945), who already stated that he "investigates synaesthesia first and foremost as a linguistic-semantic problem" (p. 812), in contrast to canonical synesthesia.

Some researchers studying sensory language invoke a strong connection between synesthetic metaphors and perceptual sensations that are akin to canonical synesthesia. For example, even though Holz (2007) says that "we have...to distinguish between a neuropsychological and a linguistic phenomenon," he also characterizes what he calls "linguistic synesthesia" as follows: "We may talk of a *verbal simulation of synesthetic perception* or of a *linguistic creation of cross-modality illusions*" (p. 193, emphasis in original). This appears to indicate that Holz believes synesthetic metaphors may trigger actual synesthetic perceptions ("*illusions*"; see also Mendelson, 1984). Anaki and Henik (2017) argue that "the processes that characterize synesthesia mirror those in metaphor processing" (p. 142).

More generally, many researchers studying crossmodal uses of sensory language try to ground linguistic patterns in language-external perceptual phenomena. Researchers have sought to explain metaphorical asymmetries (see Chapters 8–9) with recourse to evolutionary asymmetries between the senses (albeit tentatively, in Williams, 1976); differences in the diffuseness or stability of different kinds of sensory impressions (e.g., Tsur, 2012); or differences in the concreteness, salience, or accessibility of particular sensory experiences (e.g., Shen, 1997; Shen & Aisenman, 2008; Shen & Cohen, 1998; Shen & Gadir, 2009; Shen & Gil, 2007). Given that extra-linguistic perceptual facts are frequently invoked to explain linguistic patterns, it becomes important to specify what exactly is meant by the term "synesthesia" in a psychological sense, and whether this is appropriate terminology to describe linguistic expressions at all. Thus, we may ask: Is canonical synesthesia really the appropriate reference concept for expressions such as *sweet fragrance* and *rough sound*?

In the following sections, I will present four arguments against the use of the term "synesthesia" in this domain. My argument will be based on the rarity of canonical synesthesia (Chapter 6.4.1), the types of cross-sensory connections commonly attested in synesthesia (Chapter 6.4.2), the degree of voluntary control (Chapter 6.4.3), and the lack of empirical evidence for a connection between synesthesia and metaphor (Chapter 6.4.4).

6.4.1. The prevalence criterion

Estimates about the prevalence of canonical synesthesia vary widely, depending on assessment criteria and on which type of synesthesia is being investigated. However, whichever assessment criteria are used, the figures generally suggest rarity. Ramachandran and Hubbard (2001) discuss estimates ranging from 1 in 20 people having synaesthesia to 1 in 20,000. Sagiv and Ward (2006) state that about 1 in 20 people have synesthesia; Cytowic (2002) says that it is 1 in 25,000. Simner and colleagues (2006) used stricter test batteries and tested 500 university participants at the University of Edinburgh. Her team found 22 synesthetes in this group (4%). Simner, Harrold, Creed, Monro, and Foulkes (2008) report that out of 615 children aged 6 to 7 from UK primary schools, only 1.3% had grapheme-color synesthesia (seeing letters in colors). Whichever estimate we choose, synesthesia is a rare phenomenon, and some even consider its rarity as definitional (see discussion in Deroy & Spence, 2013).

This is in stark contrast to synesthetic metaphors, which have been argued to *not* be rare. Engstrom (1946) argues for their prevalence across many stylistic traditions in poetry and literature. Most researchers assume that the poetic version of synesthetic metaphors forms a continuum with such everyday expressions as *sweet fragrance* and *rough smell* (see discussion in O'Malley, 1957, p. 397–398). Engstrom (1946, p. 10) says that "our daily speech is full of synaesthetic expressions" and cites relatively mundane examples such as *heavy perfumes* and *piercing cries*. Marks (1982b, p. 15) says that "synesthetic combinations of words are much more common than most of us recognize" (see also Plümacher, 2007, p. 64; Strik Lievers, 2017, p. 84), with Whitney (1952, p. 444) calling synesthetic metaphors "an accepted and generally unnoticed part of our general vocabulary."

To demonstrate that those expressions analyzed as synesthetic metaphors are indeed quite common, one may look at the adjective–noun pair *sweet smell,* which combines a taste word with a smell word. This expression occurs 262 times in the Corpus of Contemporary American English (Davies, 2008), making it even more frequent than some corresponding literal expressions, such as *sweet taste* (105 tokens) and *sweet flavor* (101 tokens).

6.4.2. Different mappings

What specific crossmodal combinations are attested among those who qualify as synesthetes? There are many types of synesthesias that involve color, as with a case of a synesthete who experienced strong color sensations when in pain (discussed by Martino and Marks, 2001) and another synesthete reporting color sensations when experiencing bitter, sour, and sweet tastes (Engstrom, 1946, p. 7). Ramachandran and Hubbard (2001) discuss grapheme-color synesthesia as one of the most widespread forms, as does Cytowic and Eagleman (2009, p. 24). According to Simner et al. (2006), colored days of the week may be the most common form of synesthesia.

Novich, Cheng, and Eagleman (2011) provide perhaps the most extensive survey of different synesthesia types. They analyzed survey data from more than 19,000 participants and concluded that there is statistical support for five distinct subgroups of canonical synesthesia:

(a) Colored sequences synesthetes, where color sensations are triggered by ordinal sequences, such as letters or days of the weeks.
(b) Musical color synesthetes, who experience color sensations when exposed to particular types of music or sound.
(c) Colored sensation synesthesia, where color sensations are triggered by such things as touch sensations or when particular emotions are felt.

(d) Non-visual synesthetes, where a smell, sound, touch or taste is triggered (e.g., sound to taste or sight to smell).
(e) Spatial sequence synesthetes, who perceive sequences as being spatially extended, for example with numbers at certain spatial locations.

The evidence presented by Novich et al. (2011) highlights the many differences between crossmodal connections involved in canonical synesthesia versus those involved in synesthetic metaphor. Whereas touch and sound are reported to be the most frequent source and target domains in the linguistic domain (Day, 1996; Strik Lievers, 2015; Ullmann, 1959; Williams, 1976), very few canonical synesthesias involve touch, and many involve color. There are also many synesthesias involving numbers and letters that have no reflection in the linguistic expressions used by the general population.

Moreover, linguistic expressions that combine words appearing to stem from different sensory modalities generally involve modalities that are perceptually and environmentally coupled, an idea for which new evidence will be provided in Chapter 14 (see also Ronga, 2016). In contrast, canonical synesthesia is generally thought to involve inducer-concurrent pairings that are *not* based on environmental correlates (see Deroy & Spence, 2013, p. 652).

One has to further distinguish between the general type of mapping (such as colored hearing) and the specific trigger-concurrent pairings (i.e., which specific tone goes together with which specific color sensation in colored hearing). It has been reported that even synesthetes growing up in the same family can have very different pairings. For example, in grapheme–color synesthesia, the letter "R" may be blue for one synesthete but red for another (Cytowic & Eagleman, 2009). Although some regularities have been found across synesthetes, specific trigger-concurrent pairings are generally idiosyncratic

(Deroy & Spence, 2013, p. 647). This, too, is different from metaphor, where expressions such as *sweet smell* are used by hundreds of people.

6.4.3. Deliberate versus involuntary mappings

Several authors have tried to differentiate synesthetic metaphors from canonical synesthesia with respect to whether crossmodal associations are voluntary or not. Canonical synesthesia is commonly described as an automatic inducer-concurrent pairing that cannot be consciously altered. Most descriptions of synesthesia see automaticity as a defining feature of this phenomenon. Some synesthetes even describe their synesthesias as distracting. The same way that people cannot voluntarily stop seeing color, a synesthete cannot stop his or her concurrent perceptions. Some have claimed this to be different from synesthetic metaphor, which O'Malley (1957, p. 393) says "may imply a conscious, deliberate comparison of various sense qualities." Similarly, Ronga et al. (2012, p. 139), reiterating claims made by Cytowic and Eagleman (2009, p. 172), say that in contrast to canonical synesthesia, which is "completely automatic," synesthetic metaphors "require the voluntary association between words belonging to two different sensory domains." Tsur (2007, p. 49) contrasts canonical synesthesia with literary "synesthesia" by saying that the latter "leaves room for great flexibility and creativity," whereas the former is "involuntary and rigidly predictable."

While language use may be comparatively more deliberated and voluntary than the perception of synesthetic concurrents, the criterion of voluntariness does not allow a clear-cut distinction between the two phenomena. Many theories of language use, including those within cognitive linguistics, would maintain that most of the cognitive machinery that governs language use is outside the purview of voluntary and conscious control. Although it may be the case that a speaker can *choose* to describe a smell as *sweet* or not, the fact that this word came to mind at all is due to subconscious processes.

6.4.4. No evidence for a connection

In their seminal paper on synesthesia, Ramachandran and Hubbard (2001) call for investigating the connection between synesthesia and metaphor. However, so far, there is little to no empirical data that connects the two phenomena. In fact, it is not even clear what exactly one would predict if there was a strong connection between synesthesia and those expressions analyzed as synesthetic metaphors. Do we predict that synesthetes use more metaphors, or use metaphors differently? Do we predict that the expressions commonly used by speakers mirror common forms of canonical synesthesia?

Of course, metaphor and synesthesia are superficially connected in the following way: When synesthetes talk about their own synesthetic perceptions, such as describing pain as *yellow* (Martino & Marks, 2001), Wednesday as *indigo blue* (Eagleman & Cytowic, 2009), or the sound of a voice as *azure* (Engstrom, 1946, p. 6), this appears metaphorical to others who do not share the same synesthetic perceptions. For the synesthete, such descriptions are, in fact, literal. Such expressions, however, bear little resemblance to the expressions that are generally discussed in the literature on synesthetic metaphors, such as *smooth melody* and *sharp sound*.

As far as I know, the only evidence for a connection between canonical synesthesia and synesthetic metaphors is provided by a series of experiments by Marks (1974, 1975), who found that associations between pitch and brightness reported for synesthetes can also be found as crossmodal correspondences in nonsynesthetes; furthermore, English speakers use expressions such as *bright sound* and *dark sound* (see also discussion in Anaki & Henik, 2017). More research like this is necessary to allow the conclusion that canonical synesthesia and synesthetic metaphors are indeed related. However, even in the case of these pitch–brightness associations, synesthetic metaphors may be more strongly related to the widely shared crossmodal correspondences than to canonical synesthesia per se.

6.5. Summary of differences

Table 1 contrasts canonical synesthesia from synesthetic metaphors along several of the dimensions just discussed.

Table 1

Differences between canonical synesthesia and synesthetic metaphors	**Canonical synesthesia**	**Synesthetic metaphors**
Prevalence	rare	ubiquitous
Types of mappings	unassociated modalities	environmentally coupled modalities
Specific mappings	idiosyncratic	widely shared, highly conventionalized
Sensation	vivid, automatic, and conscious	no evidence for vivid and conscious perception
Voluntary control	comparatively little control	comparatively more control

Table 1 highlights that there are more differences than similarities between canonical synesthesia and synesthetic metaphors. The similarities are, if at all, very superficial: Both phenomena involve some association between different senses. Otherwise, the marriage of the terminology of perceptual "synesthesia" and linguistic "synesthesia" is an uneasy one.

Perhaps we should speak of "crossmodal metaphors" rather than "synesthetic metaphors." Crossmodal correspondences may involve environmentally coupled modalities, are not necessarily perceived in a conscious fashion, and are widely shared in the general population. All of these features apply better to cases such as *sweet smell* and *rough sound* than the notion of canonical synesthesia.

In a discussion of synesthetic metaphors from 1957, O'Malley already noted that the contrast between crossmodal correspondences ("intersense analogies") and synesthetic metaphors is less stark than that between canonical synesthesia and these expressions. This argument was based on frequency: "Intersense analogy is ancient and theoretically accessible to the experience of all normal persons. Clinical synesthesia, on the other hand, has to do with abnormal or eccentric experience of various."

In fact, some researchers have proposed that crossmodal correspondences motivate crossmodal uses of sensory adjectives. For example, the fact that people reliably associate tones of a certain pitch or loudness with particular luminances (Marks & Stevens, 1966; Stevens & Marks, 1965) may motivate such expressions as *bright sound* and *dark sound* (Marks, 1982a, 1982b).

This chapter has argued that the terminology of synesthesia has overstayed its welcome in metaphor research. I believe that adopting the terminology of synesthesia may have biased the research community in a certain way (for similar arguments, see Deroy & Spence, 2013). First, loose references to synesthesia as a mere union of the senses may have stopped linguists from fleshing out the precise mechanisms that characterize the connection between language and perception. Second, references to synesthesia may have biased researchers toward trying to find perceptual explanations for the corresponding expressions at the expense of considering language-internal explanations, such as the role of word frequency or differential lexicalization of the senses (see Chapter 15). Seeking perceptual explanations for linguistic regularities is a useful endeavor that is much in line with the Embodied Lexicon Hypothesis presented in Chapter 5. However, such endeavors should not detract from testing other, non-perceptual explanations. In general, the field of metaphor research is not helped by vaguely alluding to synesthesia.

To conclude: Synesthetic metaphors are not synesthetic. In the next chapter , I will argue that they are not metaphorical either.

Chapter 7. Synesthetic metaphors are not metaphorical

7.1. Introduction

The last chapter argued that synesthetic metaphors are not synesthetic. This chapter argues that they are not metaphorical. Strik Lievers (2017) states that it cannot be taken for granted that synesthetic metaphors are, in fact, metaphors. Following her lead, this chapter explores how expressions such as *sweet smell* and *sweet melody* fit into linguistic theory, and whether they are best characterized as metaphorical, metonymical, or literal language use.

The core question of this chapter is: What is the linguistic status of synesthetic metaphors? Do they fall within the domain of literal language use or figurative language use? If they are figurative, what type of trope are they? Chapter 7.2 begins by reviewing the basics of metaphor theory. Even though I will eventually argue that conceptual metaphor theory (Lakoff & Johnson, 1980) is the wrong theory to apply to synesthetic metaphors, a detailed discussion of metaphor theory is necessary to introduce several important distinctions. After reviewing conceptual metaphor theory, I will look at different avenues for fitting synesthetic metaphors into the cognitive linguistic framework (Chapter 7.3). Then, in Chapter 7.4, I argue for a literal analysis of expressions such as *sweet smell*— that is, an analysis where no metaphorical mappings are posited. Finally, I will argue that an evaluation-based theory of metaphor use allows the literal analysis to cover cases that appear, at first sight, to be more genuinely metaphorical (Chapter 7.5).

7.2. Conceptual metaphor theory

Probably the single most influential publication in metaphor theory was Lakoff and Johnson's (1980) *Metaphors We Live By*. This book is commonly credited with ushering in modern metaphor research, and it has created a theoretical framework that is now frequently called "conceptual metaphor theory" (Gibbs, 1994; Kövecses, 2002; Lakoff, 1987; Lakoff & Johnson, 1980, 1999), which is now considered a core component of the larger framework of cognitive linguistics (Croft & Cruse, 2004; Evans & Green, 2006). Grady, Oakley, and Coulson (1999, p. 101) say that within cognitive linguistics, conceptual metaphor is "one of the central areas of research," which is largely due to the success of *Metaphors We Live By*.

Compared to earlier views, several key innovations characterize the conceptual metaphor theory view of metaphor. First, metaphor is seen as not being confined to poetry and fanciful language use; instead it characterizes everyday language use. In fact, some researchers have estimated that up to 10% or up to 30% of words are used metaphorically, depending on the type of discourse analyzed (Pérez-Sobrino & Julich, 2014; Pragglejaz Group, 2007). The view that metaphor is not just for rhetorical embellishment existed before the advent of conceptual metaphor theory (e.g., Ortony, 1975), but Lakoff and Johnson's publication helped solidify this view. Second, within conceptual metaphor theory, metaphor is seen as a matter of thought, with language only reflecting underlying conceptual mappings. As expressed by Grady (1997, p. 281), "metaphors are based on concepts, not words." Third, Lakoff and Johnson's (1980) theory, especially later versions of it (e.g., Lakoff & Johnson, 1999), emphasized the role of embodiment more strongly. For example, as will be discussed below, conceptual metaphor theory states that many metaphors are motivated through embodied interactions with the physical world.

The empirical evidence for conceptual metaphor theory is strong. Lakoff and Johnson (1999) discuss it as one of the prime cases of "converging evidence" in cognitive linguistics, a theory that is supported by diverse strands of evidence (see Chapter 10). Whereas early opponents of the theory criticized that the evidence for conceptual metaphors was mostly linguistic (Murphy, 1996, 1997), there now is a wealth of evidence that gets at the conceptual nature of metaphor, including evidence from metaphors in gesture (Casasanto & Jasmin, 2010; Cienki & Müller, 2008; Walker & Cooperrider, 2016; Winter, Perlman, & Matlock, 2014), and from metaphors in pictures, adverts, and movies (Forceville, 2006, 2008; Forceville & Urios-Aparisi, 2009; Ortiz, 2011; Pérez-Sobrino, 2016; Winter, 2014). Moreover, there is a plethora of experiments supporting conceptual metaphor theory that are either completely or partially nonlinguistic (for reviews, see Casasanto, 2014, 2017; Landau, Meier, & Keefer, 2010; Winter et al., 2015b; Winter & Matlock, 2017).

As an example of a metaphor that has received considerable experimental support, consider INTIMACY IS WARMTH (metaphors are commonly presented in capitals). This metaphor is linguistically reflected in such expressions as *She has a warm personality*. Experiments have shown that if participants hold a warm cup, this induces positive social feelings (Williams & Bargh, 2008; see also IJzerman & Semin, 2009). On the other hand, experiments have shown that having positive social feelings makes people think that a room's temperature is warmer (Zhong & Leonardelli, 2008). These experiments provide nonlinguistic evidence for the idea that the metaphor INTIMACY IS WARMTH is based on an underlying conceptual mapping.

Let us have a look at some more precise definitions of metaphor. Whereas Kövecses (2002, p. 4) states that "in the cognitive linguistic view, metaphor is defined as understanding one conceptual domain in terms of another conceptual domain," Dancygier and Sweetser (2014) provide the following, more detailed, definition:

> A **conceptual metaphor** is a *unidirectional mapping* projecting conceptual material from one structured domain..., called the *source domain*, to another one, called the *target domain*... (p. 14, emphasis in original)

The source domain is generally thought to be more concrete and the target domain to be more abstract, although others have used familiarity, clarity, stability, accessibility, and frequency as ways of characterizing the difference between source and target domains (see e.g., Wolff & Gentner, 2011). It is often not precisely specified what exactly is meant by concreteness (Dunn, 2015), but for present purposes it suffices to characterize concreteness as "the degree to which the concept denoted by a word refers to a perceptible entity" (Brysbaert et al., 2014, p. 904). The fact that metaphors involve mappings from concrete sources to abstract targets is often treated as a definitional property of metaphors (see, e.g., Lakoff & Johnson, 1980, p. 207; Kövecses, 2002, p. 6).

Some researchers have criticized a simple notion of concreteness and reconceptualized metaphors as involving mappings from intersubjectively accessible and sharable domains to relatively more subjective and less publically accessible domains. Dancygier and Sweetser (2014) give the MORAL ACCOUNTING METAPHOR (*You owe me, I'm deep in your debt, How can I repay you for your help?*) as an example of this principle. Accounting is arguably more intersubjectively verifiable than is morality, a comparatively more subjective notion. The intersubjective accessibility of source domains is also thought to explain why physical perception is such a common source domain, such as when verbs of visual perception are metaphorically extended to indicate knowledge states—for example, when English speakers say *I see* to mean 'I understand' (Matlock, 1989; Sweetser, 1990). The objects of visual perception are generally accessible to anybody who is present in a given situation (for a discussion, see San Roque et al., 2015), whereas mental states are by definition internal and hence inaccessible to others.

Notice that the above definition by Dancygier and Sweetser (2014) specifies a “unidirectional mapping.” Many researchers take metaphors to be mapping that are in some form unidirectional, asymmetrical, or nonreversible (e.g., Kövecses, 2002, p. 6). For instance, English speakers may more frequently talk about time in terms of space (*Christmas is coming, We are approaching Halloween*) rather than the other way around (see discussion in Casasanto & Boroditsky, 2008). Within conceptual metaphor theory at least, asymmetry (or unidirectionality) is taken to be definitional of metaphor, although there are researchers who disagree with this view, especially metaphor researchers who do not align as strongly with conceptual metaphor theory (e.g., Anaki & Henik, 2017; Campbell & Katz, 2006; Katz & Al-Azary, 2017).

Unfortunately, as pointed out by Bottini and Casasanto (2010, p. 1353), Winter et al. (2015b), and Shen and Porat (2017), metaphor research often glosses over the important distinction between unidirectionality and asymmetry. Unidirectionality says that for two domains A and B, B is exclusively understood or talked about in terms of A, *never* the reverse. On the other hand, asymmetry states that two domains are bidirectionally associated with each other, but the influence of the source domain onto the target domain is stronger than the other way around.

Lee and Schwarz (2012) report an experiment testing the conceptual metaphor lying behind such expressions as *This smells fishy*, where reference to smell is used to indicate suspicion. They find that fishy smells make people more suspicious, but they also find that suspicious thoughts lead to heightened detection of fishing smells. Based on this, they argue for the bidirectionality of metaphor (see also IJzerman & Koole, 2011; Porat & Shen, 2017). However, the experiments do not allow a detailed look the asymmetry of the two domains (i.e., whether fishy smells trigger suspicious thoughts more strongly than the reverse).

Another example of confusing bidirectionality and asymmetry is provided by Winter and Matlock (2013), who test the metaphor SIMILARITY IS PROXIMITY (*These two views are close*) and show that spatial proximity influences perceived similarity, and perceived similarity influences spatial placements. The experiments do not allow for estimating whether the experimental effects of A > B are stronger than the effects of B > A, which means that the experiments have nothing to say about asymmetry. To establish genuine asymmetry over and above bidirectionality, controlled experimental conditions are necessary, where the two domains are on relatively equal grounds and can be compared directly within the same task (see, e.g., Casasanto & Boroditsky, 2008, for an attempt to investigate SPACE IS TIME metaphors). The topic of asymmetry versus unidirectionality will recur in Chapter 8.

7.2.1. Primary metaphor

Contemporary conceptual metaphor theory recognizes the existence of primary metaphors as a distinct subcategory of conceptual metaphors (Grady, 1997, 1999, 2005; Lakoff & Johnson, 1999). Primary metaphors are those metaphors that are believed to stem from embodied correlations in the environment (Casasanto, 2014, 2017; Grady, 1997, 1999; Lakoff, 1987; Winter & Matlock, 2017). For example, in our world, there is a positive correlation between verticality and quantity, as stated by Lakoff (1987):

> Whenever we add more of a substance—say, water to a glass—the level goes up. When we add more objects to a pile, the level rises. Remove objects from the pile or water from the glass, the level goes down. The

correlation is over-whelming: more correlates with up, but less correlates with down. (p. 276)

Constantly experiencing this correlation is thought to lead to the primary metaphor MORE IS UP, reflected in such expressions as *high prices*, *low prices*, and *rising taxes*. Other examples of primary metaphors include MORE IS BIGGER/SIZE IS QUANTITY (*large sum, tiny number*) and SOCIAL DISTANCE IS PHYSICAL DISTANCE/INTIMACY IS CLOSENESS (*we are very close, they drifted apart*). All of these are supposed to stem from embodied correlations—that is, persistent associations of the source and target domain in our everyday interactions (for reviews regarding the embodied nature of these metaphors, see Casasanto, 2014, 2017; Winter & Matlock, 2017). Their embodied nature is thought to make primary metaphors universal (see Grady, 1997, p. 288); that is, all cultures should show evidence for these metaphors, at least in thought (they may not be verbalized). An important aspect to which I will return below is that in primary metaphor, the source is supposed to be a concrete domain, which can be experienced through the senses, and the target is supposed to be more abstract, not as accessible to direct sensory perception (see Grady, 2005, pp. 1605–1606).

Primary metaphors can be contrasted with complex metaphors or compound metaphors, of which THEORIES ARE BUILDINGS is a prime example (Grady, 1997). Humans do not experience a strong correlation between theories and buildings. However, the metaphor rests on other metaphors such as COMPLEX ORGANIZATION IS PHYSICAL STRUCTURE and PERSISTENCE IS REMAINING ERECT (Grady, 1997, p. 273), which in turn may be motivated through embodied correlations. In contrast to primary metaphors, complex metaphors are relatively more culture-specific. In the case of THEORIES ARE BUILDINGS, a culture needs to have a concept of theories as well as a concept of buildings.

7.2.2. Metonymy

Metaphor needs to be distinguished from metonymy. Kövecses (2002, p. 144) says that "in metonymy we use one entity, or thing..., to indicate, or to provide mental access to, another entity." Metonymy is most often explained by comparison to metaphor, as Gibbs (1994) does in the following quote:

> Whilst metaphor is a process by which one domain of experience is used to refer to another unrelated domain of experience, metonymy is a process by which one aspect of a domain of experience is used to refer to another aspect of the same domain of experience. (p. 13)

For example, a speaker may say *I read Shakespeare* where she uses the name of the author to 'stand in' for his works. This expression is figurative because one cannot literally read Shakespeare, the author. The expression *I read Shakespeare* can be analyzed as involving an AUTHOR STANDS FOR AUTHOR'S WORK metonymy. With metonymy, there generally is some form of connection between the source and the target—a spatial or temporal contiguity, or a causal relationship, as in the case of an author producing a particular piece of work. Thus, whereas metaphor is generally taken to describe mappings between quite different semantic domains, metonymy is understood to involve within-domain mappings (see, e.g., Littlemore, 2015). Pérez-Sobrino (2017, Chapter 1) provides a useful analogy for distinguishing metaphor and metonymy. She likens metonymies to icebergs, where the tip of the iceberg visible above sea level indexes the presence of a larger mass of ice that is hidden beneath sea level. In contrast, she likens metaphors to bridges, as metaphors connect between different domains.

Recently, a number of scholars have begun to analyze primary metaphors as having a metonymical core (e.g., Barcelona, 2003). For example, Kövecses (2013) discussed the primary metaphor SADNESS IS DOWN/HAPPINESS IS UP, reflected in such expressions as *I'm feeling down today* and *She's having a low day*. This is a primary metaphor, posited to stem from the correlation of low vertical position with negative feelings. And indeed, people interpret downwards-oriented head posture and gaze as indicators of sadness (e.g., Coulson, 2004). The mapping between sadness and low vertical position is supposed to be motivated by contiguity. In a concrete situation, a speaker can talk about a person's low body position to index their "low" spirit because there is a causal relationship between the two (causal contiguity). Thus, Kövecses (2013) proposes that DOWN STANDS FOR SADNESS lies at the core of the primary metaphor SADNESS IS DOWN. A similar reasoning applies to MORE IS UP. In a concrete situation, such as when cookies literally pile up, there is causal contiguity (MORE leads to UP) and spatial contiguity (both MORE and UP are co-present). In such a case, using the expression *high number* can be seen as reflecting a metonymy UP STANDS FOR MORE. However, when the quantity referred to is relatively more abstract, such as taxes or interest, then the metonymical core of the primary metaphor is less in focus.

7.3. What are synesthetic metaphors?

Having reviewed different types of metaphors, we are now in a position to re-evaluate the linguistic status of synesthetic metaphors. Strik Lievers (2016, p. 43) says that the different uses of "synesthesia" refer to "at least partially different" phenomena, including different linguistic phenomena. Expanding upon the discussion in Strik Lievers (2017), the following possible linguistic analyses have to be discussed:[1]

(a) Synesthetic metaphors are conceptual metaphors
(b) Synesthetic metaphors are primary metaphors
(c) Synesthetic metaphors are metonymies
(d) Synesthetic metaphors are literal expressions

It seems that the first theoretical construct, conceptual metaphor, is not appropriate for those expressions commonly discussed in the literature on synesthetic metaphor. For conceptual metaphors, there is generally an asymmetry in the concreteness of the source and target domain, as when English speakers talk about abstract theories in terms of buildings. In synesthetic metaphors, both source and target are sensory (e.g., Shen, 2008, p. 302; Strik Lievers, 2016, p. 46; though see Petersen et al., 2008; Werning et al., 2006). Moreover, crossmodal uses of sensory adjectives do not participate in the rich inferential structures and many-to-many mappings that are the hallmark of such conceptual metaphors as THEORIES ARE BUILDINGS, where the premises and assumptions are the *foundation* of a theory, the major argument is its *framework*, a theoretician is an *architect*, and debunking a theory is equivalent to its *collapse*, et cetera (see Grady, 1997, p. 269).

It is possible that synesthetic metaphors are primary metaphors (Sullivan & Jiang, 2013). As emphasized before, expressions such as *sweet smell* often appear to involve highly interconnected sensory modalities (see also Chapter 14). In fact, several authors have pointed out that at least for some of these expressions, environmental and psychological contiguity relations play a role (Dirven, 1985; Marks, 1978, Chapter 8; Nakamura, Sakamoto, & Utsumi, 2010; Shibuya & Nozawa, 2003; Shibuya, Nozawa, & Kanamaru, 2007; Sullivan & Jiang, 2013; but see Taylor, 1995, pp. 139–140). Ronga (2016, p. 57), for example, says that crossmodal expressions involving taste "seem to reproduce the complex set of experiences that happen in the mouth." Similarly, expressions as *warm color* and *cool color* are presumably based on learned associations between color and temperature (e.g., ice and cool lakes are generally blue; fire and flames are relatively more red and yellow) (cf. Dirven, 1985; Marks, 1978, Chapter 8; see also Cacciari, 2008, pp. 429–430). Sullivan and Jiang (2013) and Shibuya et al. (2007) similarly argue that linguistic expressions that combine sight and touch (*rough color, dull color*) involve highly associated senses, not very dissimilar senses. If one is considering the role of learning through embodied experience, cases such as *warm color* appear to be similar to prototypical primary metaphors such as MORE IS UP (*high number*).

However, the concept of primary metaphor does not apply fully to the concept of synesthetic metaphor: Grady (2005, pp. 1605–1606) stresses that the target domain of primary metaphors needs to be non-sensory. As was emphasized above, synesthetic metaphors are most commonly defined as involving sources and targets that are both sensory. Thus, there is no apparent asymmetry in concreteness which would qualify synesthetic metaphors as primary metaphors. Strik Lievers (2017, p. 97) says that crossmodal uses of sensory words "may be distinguished from other metaphors because the conflicting concepts are both sensory, referring to two conceptually separate senses." Here, I would take her characterization to say that synesthetic metaphors are neither good cases of conceptual metaphors, nor good cases of primary metaphors.

A metonymical analysis may be better-suited for dealing with some cases of synesthetic metaphors, as several researchers have proposed (such as Barcelona, 2008). For instance, rather than *sweet fragrance* expressing an underlying SMELL IS TASTE metaphor, it may express a TASTE STANDS FOR SMELL metonymy. In this case, taste and smell are seen as part of the same conceptual domain, but English speakers use one sub-domain, taste, to stand in for the other related sub-domain, smell.

I am much in favor of a metonymical analysis. At least compared to a metaphor-based analysis, a metonymical analysis recognizes that sources and targets are highly associated—so much, in fact, they can be considered part of the same semantic domain. This is in line with the evidence that will be presented in Chapter 14: Sensory words tend to combine with words of their own modality, and if they do not, they tend to combine with words of highly similar modalities, such as in the case of taste and smell. A metonymical analysis recognizes this affinity and states that speakers can use one aspect of a correlated experience to stand in for another.

However, a metonymical analysis still has the problem of positing two *distinct* subaspects within the same domain. For example, within the joint domain of taste and smell, a metonymical analysis would say that speakers use taste to 'stand for' for smell. This assumes that taste and smell are separate (but correlated) qualities.

A comparison to a prototypical metonymy is insightful here. In the LOCATION STANDS FOR INSTITUTION metonymy, exemplified by such statements as *The White House pardoned Joe Arpaio*, the location sense of *White House* is clearly separate from the institution sense. Can we say the same thing about an expression such as *sweet smell*? Are taste and smell separate enough from each other to warrant a 'stands for' relationship that connects two separate senses? Or is no such relation needed because the word *sweet* is, in fact, as much a smell word as it is a taste word?

7.4. The extent of the literal

7.4.1. The role of multisensory perception

Dancygier and Sweetser (2014) remind us that "thinking about figurative expressions requires that we develop hypotheses about how words can provide access to concepts which are not literally associated with them" (p. 13). This is directly related to the crux of this chapter: Both metaphor- and metonymy-based accounts assume that there are additional concepts "which are not literally associated" with a sensory word's core meaning. For example, analyzing *sweet fragrance* as being a taste-to-smell metaphor (Shen & Gil, 2007) rests on the assumption that 'smell' is not already part of the meaning of *sweet*. In general, positing a mapping between two entities requires two separate entities to be mapped onto each other. However, this assumption may be at odds with the intense multisensoriality of sensory words.

My argument in this section will revisit several points made by Marina Rakova (2003) in her book *The Extent of the Literal*, where she argues that sensory adjectives such as *sweet* and *harsh* have rich, highly supramodal semantic content with a much broader denotational range than is commonly assumed (see also Marks, 1982a, p. 192; Paradis & Eeg-Olofsson, 2013). In terms of lexical semantics, her work implies that the two expressions *sweet taste* and *sweet smell* do not reflect two separate polysemous senses, but two contextually modified applications of the same underlying meaning (vagueness or "lack of specification"; see Cruse, 1986; Zwicky & Sadock, 1975). Thus, Rakova's (2003) proposal amounts to saying that the word *sweet* has just one meaning that denotes both taste and smell.

Rakova (2003) exemplifies her approach with the sensory adjective *hot*. This word appears to have two distinct meanings, one referring to temperature, as in *hot stove*, and another one referring to spiciness, as in *hot chili paste*. At first sight, temperature and spiciness appear to be two quite dissimilar perceptual qualities. However, Rakova (2003) uses neurophysiological evidence to show that the two perceptual meanings actually correspond to the same underlying neural system. The evidence has to do with nociception, our sense of pain. One pain receptor, the so-called “TRPV1” receptor responds to both noxious heat (with a thermal activation threshold of ~43 °C) and to capsaicin, a molecule that is present in chili peppers (e.g., Basbaum, Bautista, Scherrer, & Julius, 2009; Julius & Basbaum, 2001). Physiological responses to hot food (in terms of temperature) and to spicy food are moreover experienced to be phenomenologically similar, including sweat and the sensation of heat on the skin. Rakova (2003) then surmises that the concept expressed by the word *hot* denotes both spiciness and noxious heat sensations.

One easy misunderstanding of Rakova’s account is to see it as requiring us to posit that words directly refer to underlying perceptual brain states without conceptual mediation. Even though some do indeed assume such a direct mapping from words to perception (see Fahle, 2007, p. 35), it has been argued that language does not refer to perception directly, but mediated through intervening cognitive steps (see, e.g., Miller & Johnson-Laird, 1976).[2] In fact, according to the discussion of “ineffabilities” in Chapter 4, this is a necessary component of perceptual language, which involves such conceptual procedures as categorization (grouping particular experiences together). When Rakova (2003, p. 42) says that the spicy–hot association “is grounded in the molecular constitution of our pain detecting mechanisms,” this does not necessitate us to posit an unmediated match between brain states and words. Rakova herself actually endorses the view of highly “supramodal” concepts (e.g., p. 142) that *result* from the underlying neural association. For instance, speakers may have a high-level supramodal concept of hotness–spiciness that results from the repeated experience of similar sensations.

If one posits that sensory words are connected to such supramodal conceptual representations in which several sensory modalities are associated, is this not the same as a primary metaphor- or metonymy-based analysis, which also emphasize embodied associations? The difference between a figurative and a literal analysis is that the latter does not require a mapping from temperature to spiciness in cases such as *hot food*. That is, English speakers do not use temperature to understand spiciness; both are recognized as equally basic. Instead, *hot stove* and *hot chili paste* are seen as contextual modulations of the same literal meaning.

One can similarly use neurophysiological evidence about other sensory systems to carry her analysis of the meaning of *hot* to other perceptual domains. The argument easily extends to taste and smell. Chapters 13 and 14 will provide new evidence for the idea that linguistically, these two domains are inseparable. Spence and Piqueras-Fiszman (2014, p. 188) cite Jean Anthelme Brillat-Savarin, who said that "smell and taste are in fact but a single sense, whose laboratory is in the mouth and whose chimney is in the nose." In Ronga's (2016) study of taste metaphors, she acknowledges that "it is not possible to disentangle" whether particular words "selectively refer to taste or olfaction" (p. 51). A similar point is made by Lehrer (1978, p. 98), who noted that "words denoting taste cannot always be separated clearly from those for feel and smell" (cf. Staniewski, 2017).

Lehrer's quote naturally leads to us to discuss the interconnection between touch and taste as another case of highly associated modalities that may motivate linguistic expressions such as *smooth taste*, *rough taste*, and *sharp taste*. Lehrer (1978, p. 119) noted: "Since we have touch receptors in our mouth, it is easy enough to understand how touch sensations could be transferred to taste." There is a whole wealth of tactile sensations felt in the mouth (Lehrer, 2009, p. 7). Our sense of touch is a quintessential component of flavor perception, participating in many behavioral interactions with taste and smell (see Auvray & Spence, 2008; Spence & Piqueras-Fiszman, 2014).

The argument for a shared neural substrate motivating the various uses of sensory words also extends to pain words. Semino (2010) discusses corpus examples of the adjective-noun phrase *sharp pain* (p. 208). In her account, the expression *sharp pain* is often best analyzed as a metonymical expression, where the cause of a pain (e.g., a sharp knife) is used to describe the evoked sensation (e.g., what one feels when being cut by a sharp knife). Furthermore, when *sharp pain* is used to describe pain that does not result from the damage of external entities (e.g., *sharp pain in my stomach*), she analyzes *sharp* as metaphorical (see also Schott, 2004). According to her, "even when pain does not directly result from tissue damage, it tends to be described metaphorically in terms of a variety of causes of physical damage" (p. 223). However, the pain evoked from external stimuli and the pain felt via interoception (i.e., the sense of the internal state of the body; see Connell, Lynott, & Banks, 2018) have common neural pathways and may ultimately be represented mentally as an emotion (see Craig, 2003). Thus, words such as *sharp* may tap into highly general sensory concepts that encompass multiple types of pain, both internal and external, as well as the resulting emotional response.

We may even extend the literal argument to expressions involving touch and sound, such as *rough sound* and *abrasive sound*. There is abundant evidence for deep interconnections between audition and touch (e.g., Guest, Catmur, Lloyd, & Spence, 2002; Jousmäki & Hari, 1998; Lederman, 1979; Levänen, Jousmäki, & Hari, 1998; Schürmann, Caetano, Jousmäki, & Hari, 2004; Suzuki, Gyoba, & Sakamoto, 2008), including single-cell recording studies showing that the macaque auditory cortex (and hence, presumably the human auditory cortex as well) has neurons that directly respond to both auditory and somatosensory stimuli (Schroeder et al., 2001). Similar to the case of taste and smell, touch and sound are much closer than a metaphor-based account would suggest, and so, too, are touch and taste. Hence, touch words such as *rough* and *abrasive* may be referring to highly general and highly supramodal concepts that are rendered specific when used in contexts such as *rough taste* and *rough sound*.

It is useful to compare the usage of words such as *rough* and *sweet* in different contexts to the usage of color words, such as *red*. There is a continuum underlying color perception, defined by the physical quantity of wavelength. Color words such as *red* demarcate particular points on this continuum. A wavelength of about 650 nanometers, for example, is associated with the label *red* in English. Small variations in this wavelength (say, 645nm and 655nm), are still within the color category denoted by the word *red*. When humans perceive such small variations in wavelength, slightly separate but largely overlapping clusters of neurons are activated. In some cases, the wavelength may even be quite different and a speaker would still label the color *red*, as in the case of a *red brick*, which in many countries (such as the US) is often more orange than red. A similar example is provided by Cacciari (2008, pp. 425–426), who mentions the fact that in a language such as English, the word *blue* is used to describe both the sky and the ocean, even though the chromatic characteristics of the corresponding percepts are quite different. Yet, in neither of these cases do we want to posit two separate senses, red_1 (for *red brick*) and red_2 (for *red rose*), or $blue_1$ (for *blue sky*) and $blue_2$ (for *blue ocean*). Instead, we recognize that the word *red* has a fuzzy boundary.[3] Given this, there is no need to invoke a mapping from one color to another color, be it metonymical or metaphorical. The expressions *red rose* and *red brick* are both analyzed as literal.

The argument then is that using *red* to talk about a *red brick* is not qualitatively different from using *sweet* to talk about *sweet smell*. And it is not qualitatively different from using *hot* to talk about *hot food* (with the meaning 'spiciness'). Just as different types of redness sensations activate slightly different but also partially overlapping neural circuits, so too do different types of multisensory taste–smell sensations and temperature–spiciness sensations. In both the color case and the taste–smell case, there are two gradations of one underlying concept, two nuances of the same meaning. Because of this, the same way that we do not consider *red brick* to involve a figurative application of the word *red*, we do not have to consider *sweet* in *sweet fragrance* to be figurative either (contrary to, e.g., Shen & Gil, 2007).

7.4.2. Categorical intuitions

On the other hand, our intuition tells us that *sweet* is a taste word. A supramodal account is not inconsistent with this intuition, as it allows for some meaning components to be more salient than others. The same way that the redness sensations denoted by *red apple* may be more prototypical than those denoted by *red brick*, the use of *sweet* in *sweet taste* may be more prototypical than the use of *sweet* in *sweet smell*.

Moreover, one has to think about from where the intuition that *sweet* is a taste word comes. Crucially, this intuition is not devoid of cultural context, even if performed by a linguist. After all, most linguistic research either explicitly or implicitly assumes the five senses folk model (Chapter 2), which is first and foremost a cultural model. Classifying the word *sweet* as a taste word is not a culture-free process; it slots the word into a particular view of the sensory world.

What is more, it should be noted that a question such as "Is *sweet* a taste word?" is a loaded question because it presumes categoricity. We similarly expect a categorical answer when someone asks "What type of sense is described by *sweet*?" Spivey (2007) discusses how particular behavioral tasks impose categoricity, which may make cognitive processes that are actually continuous seem categorical. I would argue that this is exactly what happens when researchers treat *sweet* as categorically belonging to taste. In fact, Lynott and Connell (2009) show that when people are given a continuous scale and are asked to focus on each sensory modality separately, they actually rate *sweet* to be high in both gustatory and olfactory content. Thus, when using a different task, one arrives at a different view of the word *sweet*. Connell and Lynott (2016) show that when people consciously think about sensory words, they are unaware of the multisensory nature of these words unless particular sensory modalities are highlighted individually.

Connell and Lynott (2016) also show that their multisensory measurement scale outperformed relatively more unisensory measurement scales in predicting performance in a word processing task. Similarly, in Chapters 15 and 16, I will demonstrate that continuous modality scales, rather than categorical classifications, are better at predicting linguistic patterns such as word frequency and evaluative language use. This suggests that in actual language use, categories do not matter as much as the continuity of the senses.

To conclude this section: When Strik Lievers (2017, p. 93) says that "linguistic descriptors of sensory experience...tend to be classified as pertaining to one sense or another," she is certainly right, but we have to ask the question: What compels us to classify? And who is doing the classifying? It appears to me that it is primarily linguists who do the classifying, primarily for the purpose of facilitating linguistic analyses that already assume distinct categories in adherence to the five senses folk model. When linguists classify an expression such as *sweet fragrance* as a taste-to-smell mapping, this analysis imposes categoricity and unisensoriality onto the word *sweet* in part because of the linguist's own cultural belief systems, in part because he or she is looking at the word in isolation, in part because he or she is asking a question that demands a categorical answer, and in part because the multisensory nature of sensory words is difficult to intuit (Connell & Lynott, 2016). Actual language use does not need to obey such categorization processes.

7.5. Evaluation and conceptual conflict

7.5.1. Conceptual conflict

The argument presented so far only applies to uses of sensory words that involve perceptually associated modalities. Explanations based entirely on associated modalities do not explain the full scope of what people consider synesthetic metaphors, as highlighted by Lehrer (1978):

> In the case of touch, taste, and smell, it may be that simple association will do—the association of how foods and beverages taste and fell in the mouth and how they taste and smell. However, the transfers from touch, taste, and dimension to sound and color and the transfers between sound and color would seem to be genuine synaesthetic transfers, and they call for an explanation. (p. 121)

Strik Lievers (2016, p. 45), invoking Prandi's (2012) notion of conceptual conflict, provides a useful definition of synesthetic metaphors according to which they involve a combination of sensory lexemes that "generate conflict at the conceptual level." The expression *sweet melody* would be analyzed as involving conceptual conflict because melody is an auditory construct that has no gustatory manifestation. For the argument presented so far, such examples are problematic because in contrast to cases such as *sweet smell,* it is not as easy to posit an experiential contiguity, or any strong perceptual association. Although there are interactions between taste and sound (e.g., Crisinel et al., 2012), the involved brain structures do not appear as overlapping as in the case of taste and smell, which arguably form one shared neural system. Taylor (1995, p. 139) lists synesthetic metaphors such as *loud color* as expressions that “cannot reasonably be reduced to contiguity.” Following Strik Lievers’ (2016) notion of conflict, we may say that there are conflict-involving expressions (e.g., *sweet melody*) and non-conflict-involving expressions (e.g., *sweet smell*).

My argument presented so far only applies to non-conflict-involving expressions. So, are conflict-involving expressions genuine synesthetic metaphors? Here one needs to ask: What exactly is it that is supposed to be mapped in expressions such as *sweet melody*, if anything is mapped at all? Is it the case that *sweet melody* actually involves the mapping of specific gustatory context onto an auditory concept?

7.5.2. The role of evaluation

Many researchers have argued for a role of evaluative and emotional meaning in synesthetic metaphors (e.g., Barcelona, 2003, 2008; Lehrer, 1978; Marks, 1978; Mendelson, 1984, p. 350; Shibuya & Nozawa, 2003; Shibuya et al., 2007; Tsur, 2012). Already Osgood (1963, pp. 346–347) surmised that evaluation may be what motivates metaphorical mappings, including synesthetic metaphors (see discussion in Lehrer, 1978, p. 121). In general, Lehrer (2009, Chapter 6) and others recognize that perception and evaluation are often inseparable. In the context of wine, she says, “Wine drinking is an aesthetic experience, and naturally, the evaluative dimension is important and permeates every other dimension, including the descriptive ones” (p. 7).

Lehrer (1978) provides an illustrative example of how something that appears to be synesthetic ceases to be so once one looks at it from the perspective of evaluation and emotions:

> Joseph Williams predicts that if someone were asked to run a wire brush over his hand and say whether it felt *sweet* or *sour*, he would have no difficulty judging. However, the response might not be synaesthetic but rather a transfer of the meanings 'pleasant' and 'unpleasant'. (p. 119, Footnote 12)

In fact, in perceptual science—regardless of any linguistic considerations—researchers have begun to explain certain crossmodal correspondences as being based on emotional processes as well. Palmer, Schloss, Xu, and Prado-León (2013) provide experimental evidence that associations between music and color are not based on perceptual correspondences between music and color, but on emotional correspondences (e.g., between major mode and happiness).

Within language, Barcelona (2003) and Tsur (2012) discuss the examples *loud color* and *loud perfume*, both of which appear to primarily involve the mapping of "annoyingness" onto the visual and smell domain respectively, rather than the mapping of specific auditory content. Similarly, Lehrer (1978, p. 121) says that the expression *sour note* is used "not because the note sounds as if it would taste sour," but because the feature "[Displeasing to the Senses]" is borrowed from the taste domain and applied to the description of a sound. Following the lead of these authors, my argument will be that in some cases, evaluation trumps perceptual content when it comes to word choice. That is, in such expressions as *sweet melody*, the highly evaluative meaning of *sweet* may be the dominant factor, not the specific gustatory content of the word.

Clearly, speakers consider both denotational content and evaluative content when choosing words. An extreme example for this is the word *spinster*. Purely based on its descriptive meaning, one should be able to use the word to apply to all elderly unmarried women. However, the clearly derogative meaning of the word prevents usage in most contexts, even if there is denotational fit.

Words such as *spinster* and prototypical evaluative words such as *good* and *bad* are not the only words to have evaluation as a core part of their lexical representation. Lehrer (1978, p. 121) already noted that "perhaps the basic semantic features that permeate all sensory words are those of intensity and evaluation." And indeed, as reviewed in Chapter 5, there is psycholinguistic evidence that evaluative meaning is a core part of a word's lexical representation (e.g., de Houwer & Randell, 2004; Vigliocco et al., 2009; Warriner & Kuperman, 2015). Unfortunately, the language sciences often have a denotative bias and do not consider the importance of evaluative meaning in lexical semantics and metaphor research. Yet, perceptual and emotional meaning together may explain how sensory words are used crossmodally.

The analysis presented here is actually consistent with the idea that there is a trade-off between perceptual and evaluative meaning in a word's lexical representation, as per Vigliocco et al. (2009), Kousta et al. (2011), and Meteyard et al. (2012). In fact, this trade-off between perceptual specificity and emotional involvement already came up when discussing smell terms in Chapter 4: Smell-related adjectives such as *fragrant*, *aromatic*, *pungent*, and *rancid* do not specify olfactory content in a precise manner, but they are quite clearly either positive or negative (cf. Levinson & Majid, 2014; Majid & Burenhult, 2014). Ankerstein and Pereira (2013, p. 312) talk about a similar aspect of taste language: The highly evaluative word *sweet* can describe various tastes that differ quite starkly from each other, "it does not offer a clear description of a particular taste." The visual adjective *blue* on the other hand, is much more specific, and it is also comparatively more neutral. Miller and Johnson-Laird (1976, p. 356) discuss a similar trade-off, saying that evaluative adjectives such as *good* are much more semantically dependent on their head nouns than color adjectives. A trade-off between emotional and perceptual content was also proposed in the context of synesthetic metaphors by Shibuya et al. (2007), who state that crossmodal metaphors are grounded either in a perceptual connection between two words (e.g., taste and smell), or in an emotional connection (see also Shibuya & Nozawa, 2003). Both descriptive and evaluative uses of words co-determine their use in linguistic expressions, including in linguistic expressions that are analyzed as synesthetic metaphors.

Chapter 17 will present additional empirical evidence for the emotion–perception trade-off hypothesis.

7.5.3. The metaphor way of dealing with evaluation

Whereas most corpus linguists view evaluation as a core component of language (e.g., Hunston, 2010; Thompson & Hunston, 2000), researchers in the cognitive linguistic tradition have treated evaluation and emotional meaning, either explicitly or implicitly, as something special. In particular, researchers working in this tradition have posited conceptual metaphors for cases where a concept is "mapped" across domains for mostly evaluative purposes (see analysis of taste metaphors in Sweetser, 1990).

To exemplify some of the problems inherent in a metaphor-based approach of dealing with the evaluative use of adjectives, let us have a look at Bagli's (2016) analysis of how Shakespeare uses the words *sweet*, *bitter*, *sour*, *tart*, *salt*, and *spicy* "metaphorically." Bagli finds sweetness to be the most prolific source domain (p. 149), used in expressions such as *Marry, sir, because silver hath a sweet sound* (p. 151). This usage of *sweet* in *sweet sound* is analyzed by Bagli (2016) to reflect the metaphor HEARING IS SWEET. For other expressions, he posits such metaphors as CHILDHOOD IS SWEET (*A mother only mock'd with two sweet babes*), LOVE IS SWEET (*steal love's sweet bait*), and RECOVERY IS SWEET (*The sweetest sleep*). It should be noted that many of the target domains appear to have positive connotations, which is reflected in Bagli's analysis by assuming a more general PLEASURE IS SWEET metaphor from which the other metaphors spring. For the word *bitter*, Bagli (2016) posits such metaphors as DISPLEASURE IS BITTER, EVIL IS BITTER, and SORROW IS BITTER, all of which involve negatively connoted target domains. Among the metaphors involving *sour*, Bagli lists negatively connoted metaphors such as DANGER IS SOUR, DISPLEASURE IS SOUR, and SORROW IS SOUR.

The fact that there is a consistent evaluative component running across several metaphors suggests that a simpler analysis may be possible, one which recognizes the inherent evaluative meaning of words such as *sweet* and *bitter*. Saying that *For these bitter tears* (p. 154) is SORROW IS BITTER and that *The consequence will prove as bitter* (p. 153) taps into a different DISPLEASURE IS BITTER metaphor does not appear to recognize the inherent evaluative similarity between these two uses. Moreover, in positing the conceptual metaphor DISPLEASURE IS BITTER, we state that there is a mapping between emotional valence (DISPLEASURE) and taste, even though the displeasing nature of the word *bitter* seems to be an inherent aspect of the word's meaning. If, as Dancygier and Sweetser (2014) remind us (see above), figurative language indeed provides access to what is otherwise *not* literally associated with a word, positing the metaphor DISPLEASURE IS BITTER amounts to saying that bitter sensations are not generally associated with displeasing feelings.

One can easily see that Bagli's metaphor-based analysis, which is also reflected in such approaches as Barcelona's (2003, 2008), would amount to a proliferation of metaphors. For example, *ugly* and *attractive* are both dominantly visual words according to the native speakers consulted in Lynott and Connell's (2009) rating study. When someone says *ugly smells* or *attractive sounds*, is it necessary to specify new mappings for each of these uses? What about the gustatory adjective *tasty*, which would then seem to require something like a PLEASURE IS TASTY mapping for such uses as *tasty boogaloo beats* (example from Pang & Lee, 2004)? It must be recognized that in many contexts, words such as *sweet* and *bitter* do not behave much differently from words such as *tasty*, *distasteful*, *palatable*, and *unpalatable.*

7.6. Conclusions

Synesthetic metaphors are not metaphorical after all. In the case of expressions involving no conceptual conflict, such as *sweet smell*, it appears most plausible to assume highly supramodal lexical representations, following Rakova (2003). In conflict-involving cases, such as *sweet melody*, evaluation matters relatively more strongly. Others have made conceptually similar proposals. Thus, the phenomenon of synesthetic metaphor ceases to be metaphorical for two reasons: either there is no need to posit synesthetic metaphors because the involved perceptual modalities are highly integrated, or there is no need to posit synesthetic metaphors because crossmodal uses simply follow from word-inherent evaluative meaning.

Thus, synesthetic metaphors are grounded either in perceptual association or in the evaluative meaning of words (compare Shibuya & Nozawa, 2003; Shibuya et al., 2007). This view is also consistent with the fact that emotional meaning is part of a word's lexical representation (Chapter 5) and there may be a trade-off between abstractness and emotional meaning in the lexicon (Kousta et al., 2011).

It should be emphasized that the view expressed here, although it may at first sight appear to stand against certain commonly held beliefs in cognitive linguistics, is in fact fully embodied. Our emotional response to sweet, sour, bitter, and other tastes is fully embodied. It is this hedonic response that is connoted by the corresponding words, and this subsequently affects how the words denoting these tastes—*sweet, sour, bitter*—are used in context (cf. Bagli, 2017, p. 46). The embodied account advocated for in this chapter recognizes lexical presentations come with emotional meaning, a view for which there is independent experimental evidence (see Chapter 5).

The approach advocated here is also embodied in the sense that it is compatible with the evidence for mental simulation. According to Gibbs (2006), Gibbs and Matlock (2008), Semino (2010), and many others, understanding a metaphor involves the perceptual simulation of a metaphor's source domain. This view is consistent with the large bulk of evidence for perceptual simulation stemming from both behavioral and neuroscientific studies (reviewed in Chapter 5), and it has also received direct support from studies on metaphor processing (e.g., Lacey et al., 2012; Wilson & Gibbs, 2007). In line with the Embodied Lexicon Hypothesis, I fully acknowledge the evidence for mental simulation. However, I would go one step further and argue that because the simulations themselves are already multisensory, the notion of metaphor does not need to be evoked for such expressions as *sweet smell* and *rough taste*. Thus, my account is fully embodied and fully simulation-based, but it makes do without metaphor.

In fact, following Rakova (2003), one might argue that the analysis here is even more embodied than metaphor-based or metonymy-based analyses of synesthetic metaphors. Paradoxically, it is precisely the linguistic literature that is most directly devoted to dealing with the multisensoriality of sensory words—the literature on synesthetic metaphor—that tacitly imposes unisensoriality on sensory words, a unisensoriality that is not grounded in perceptual facts. Research on synesthetic metaphor thus falls prey to the five senses folk model, which is incorporated into linguistic theorizing as if it is an established fact rather than a cultural assumption. The very notion of synesthetic metaphor stands against the continuity of the senses, because it implies mappings between discrete domains. Once we take evidence from sensory science and psycholinguistics into account, no such mappings need to be posited. Thus, although I have come from a slightly different angle, I conclude with Rakova (2003, p. 147) that "all meanings of synaesthetic adjectives are likely to be their literal meanings."

---

[1] It should be noted it is possible not all linguistic expressions that are commonly thought to be synesthetic metaphors may be accounted for by just one analysis approach. Perhaps different types of expressions require different linguistic analyses.

[2] Miller and Johnson-Laird (1976, p. vii) say that the correlations they had noted between language and perception "were mediated by an enormously complex conceptual structure... Percepts and words are merely into and out of this conceptual structure. Any theory of the relation between perception and language must necessarily be a theory of conceptual thought." Elsewhere they say "the notion that language is a process of associating vocal noises with perceptual stimuli seems too simple" (p. 177).

[3] Plümacher (2007) says that "we do not designate every shade which is visually identifiable, but confine ourselves to names with a wide range of application" (p. 61; see also Fahle, 2007, and Wyler, 2007).

Chapter 8. The hierarchy of the senses

8.1. Introduction

The last two chapters deconstructed the notion of synesthetic metaphor, arguing that synesthetic metaphors are neither synesthetic nor metaphorical. This section deals with the hierarchy of the senses, the idea that the senses can be ranked according to how frequently they are used to talk about the other senses. Ullmann (1945, 1959) proposed the following order:

(a) touch < heat < taste < smell < sound < sight

Following this hierarchy, metaphorical mappings are understood to move from left to right. Modalities to the left of the hierarchy are presumed to be more likely sources. Modalities to the right are presumed to be more likely targets. Examples of hierarchy-consistent expressions are *coarse voice* (touch-to-sound), *sharp taste* (touch-to-taste), *warm color* (heat-to-sight), and *sweet fragrance* (taste-to-smell). The hierarchy of the senses, if it is true, is striking: A complex web of inter-sensory relations (assuming five senses, at least 25 combinations), is reduced to a highly general and abstract linear hierarchy.

The first pieces of evidence for the hierarchy came from an analysis of Byron and Keats, for which Ullmann (1945) noted a higher proportion of hierarchy-consistent expressions than hierarchy-inconsistent expressions. Ullmann (1959) later confirmed the same tendencies in literary works from other languages, including French and Hungarian. A higher proportion of upwards-transfers was subsequently found in a number of Indo-European languages, including English, Italian, German, and Russian (Day, 1996; Callejas, 2001; Mendelson, 1984; Ronga et al., 2012; Strik Lievers, 2015; Williams, 1976), as well as in a number of non-Indo-European languages, including Hungarian, Chinese, Japanese, Hebrew, and Indonesian (Erzsébet, 1974; Shen, 1997; Shen & Gil, 2007; Whitney, 1952; Williams, 1976; Yu, 2003). Detailed studies of particular modalities also corroborate the notion of a hierarchy. For example, Bagli (2017) uses data from the "Mapping Metaphor with the Historical Thesaurus" project to show that the sensory modality of taste is more frequently used as a source domain than as a target domain across the history of English.

Experimental studies provide converging evidence for the hierarchy of the senses, showing that hierarchy-consistent examples such as *sweet fragrance* are more easily recalled in memory tests than hierarchy-inconsistent examples such as *fragrant sweetness* (Shen, 1997; Shen & Aisenman, 2008), and they are also judged to be more natural, accessible, and easier to interpret (Shen & Cohen, 1998; Shen & Gadir, 2009; Shinohara & Nakayama, 2011; Werning et al., 2006).

8.2. Different versions of the hierarchy

In the literature on synesthetic metaphors, several modifications of Ullmann's original hierarchy have been proposed (see Shinohara & Nakayama, 2011). One of the more complex hierarchies is presented by Williams (1976).[1] This hierarchy is shown in Figure 1 with example expressions that identify each path.

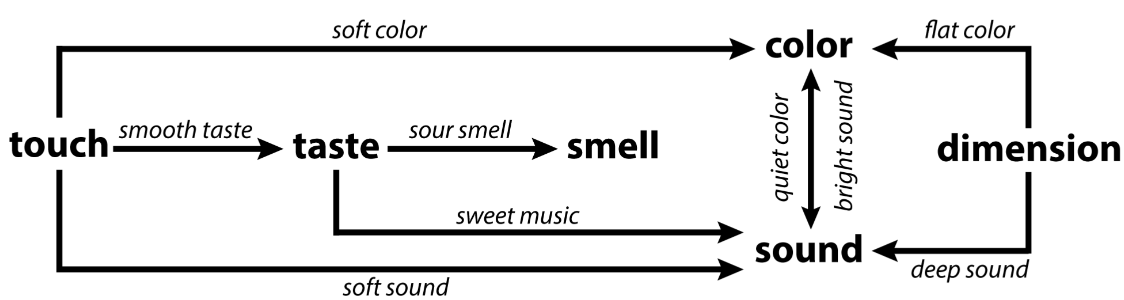


Figure 1. The sensory metaphor hierarchy according to Williams (1976, p. 463).

There are a large number of differences between the hierarchy of Williams (1976) and that of Ullmann (1959). First, the two use different sense classifications, with Williams (1976) lumping Ullmann's (1959) touch and heat categories together. This has generally been done in most studies of crossmodal expressions since then (but see Day, 1996; Ronga et al., 2012). Williams (1976) labelled the vision node "color," which de-emphasizes visual characteristics such as brightness, saturation, and opacity and spatial characteristics such as shape. However, as pointed out by Ronga et al. (2012, p. 464), among the adjectives he considered, there are many non-color-related adjectives such as *bright, brilliant*, *clear*, *dark*, *dim*, *faint*, *light*, and *vivid*. These words denote visual characteristics other than hue, and hence, we can think of the color node in his hierarchy as a visual node.

Williams (1976) captures the spatial dimension by adding a new category, dimension words, which describes such things as extent and shapes, such as the words *acute*, *big*, *small*, *deep*, *empty*, *even*, *fat*, *flat*, *low*, *high*, *hollow*, *level*, *little*, *shallow*, *thin, thick, large*, and *full* (p. 464). While it is plausible to separate these words describing common sensibles (see Chapter 4,) from other sensory words, Ronga et al. (2012) rightly points out that Williams' (1976) classification of words as dimension words lacks explicit criteria (see also discussion in Lehrer, 1978).

It must then be highlighted that Williams (1976) discusses a different set of crossmodal connections. In this model, sight and sound are bidirectionally connected, with expressions such as *quiet color* (sound-to-color) and *bright sound* (color-to-sound) exemplifying both paths. This aspect of Williams' hierarchy was preshadowed by Ullmann (1959, Chapter 5.2), who already noted that the relative positioning of the visual and auditory modality is not necessarily clear. Based on his corpus evidence, Ullmann already noted that there appears to be a bidirectionality between sound and sight, something that has been found repeatedly since then (e.g., Day, 1996; see also Marks, 1982a, 1982b). Although Ullmann's visual representation of the hierarchy in tabular format lists sound before sight, his writing suggests that the two are actually in the same position within the hierarchy. Shen and colleagues have since then also treated sound and sight as equipositional (e.g., Shen & Gil, 2007).

It should be noted that in the network of asymmetrical sense relations of Williams (1976), the domain of smell is an impasse: It borrows linguistic material from the other modalities, but it does not lend any material of its own (see also Sadamitsu, 2004). This was already present as a numerical trend in Ullmann's data, where there were few cases of smell words used in the context of sound words (Tsur, 2008, p. 288). Finally, it should be noted that Williams' (1976) hierarchy does not have a direct touch-to-smell connection (see also Lehrer, 1978).

The following simplified hierarchy is a representation of the hierarchy that most researchers in the field appear to adopt (e.g., the work of Shen and colleagues), what we might call a "consensus hierarchy" or "simplified hierarchy":

(b) touch < taste < smell < sound / sight

We should ask whether the crossmodal "transfers" specified by the hierarchy are characterized by asymmetry or unidirectionality, an important distinction discussed in Chapter 7. That is, does one sense "transfer" to another but *never* the reverse (unidirectionality), or does one sense "transfer" to another more so than the reverse (asymmetry)? Given the already established evidence, it must be the case that the hierarchy of the senses is about metaphorical asymmetry rather than unidirectionality. This follows from the fact that Ullmann

(1959), Shen (1997), Strik Lievers (2015), and many others report cases that go against the hierarchy, but these cases are much less frequent.[2]

However, once we are dealing with asymmetry rather than unidirectionality, we have to worry about the fact that the asymmetry may be stronger or weaker for particular combinations of sensory modalities. The visual representation of hierarchies such as (a), (b), and Figure 1 appears to suggest that all pathways between the senses are equally likely, despite corpus work showing that this is not the case. Among other things, several researchers (including Ullmann) found touch-to-sound "mappings" to be overrepresented (see Day, 1996; but also see Whitney, 1952). Unfortunately, many publications in this field do not allow for assessing the strength of particular crossmodal connections because only summary data is presented. For example, Shen (1997, p. 50) says that of the 130 synesthetic metaphors in his Hebrew poetry corpus, 95 matched the hierarchy (73%), a figure that does not allow assessing which particular sensory associations are more or less dominant. This tradition of reporting was started with Ullmann (1945, 1959), who only listed detailed tabulations for Keats, Byron, and Gautier, but not for any of the other poets he analyzed, such as Longfellow and Wilde.

8.3. Conclusions

In this chapter, I discussed the notion of a hierarchy of the senses, of which there are multiple different versions. This hierarchy is supposed to govern asymmetries in how sensory words are combined. How can the idea of a hierarchy discussed here be reconciled with the idea, presented in the last chapter, that synesthetic metaphors do not really involve metaphorical mappings? It has to be recognized that the very idea of a hierarchy of the senses where different modalities are lined up in a linear fashion necessitates differentiating the senses to begin with; that is, we can only talk about touch being at the lowest position in the hierarchy if we take touch to be a well-demarcated "thing" that can occupy a clear position in an ordered list, separate from the other modalities. The hierarchy of the senses thus rests on the assumption of the five senses folk model: Only by assuming five sensory modalities can we position these modalities with respect to each other. It would seem that if we dispense of the five senses folk model and the idea that synesthetic metaphors involve metaphorical mappings, the hierarchy of the senses falls apart.

The idea of specific crossmodal uses of sensory words being more frequent than others is, in fact, compatible with a more continuous view of how words are associated with sensory modalities. It is possible to maintain that *sweet* is both gustatory and olfactory while at the same time observing that it may appear to be primarily a taste word when one deliberates the word's meaning in isolation (see Chapter 7). The question then is not about whether there is a "transfer" of taste meaning to smell meaning, but whether words whose semantic prototypes are relatively more taste-related occur in linguistic contexts that are relatively more smell-related. There still can be empirically observed asymmetries in how the sensory vocabulary is deployed in language use, but these asymmetries are then about words with multiple meaning dimensions that get shifted in emphasis when they occur in specific contexts. The only thing avoided is the assumption of a discrete process of understanding one clearly delineated sensory domain in terms of another clearly delineated sensory domain. Chapters 14 and 17 will demonstrate that crossmodal language use can be analyzed continuously.

For the sake of discussion, we will temporarily accept the existing evidence for the hierarchy (but see Chapter 17 for new evidence and qualifications) so that we can move to a discussion of what *explains* the hierarchy. In the next chapter, the hierarchy of the senses will be discussed from a multicausal perspective, looking into several of the factors that may contribute to asymmetries in the usage of sensory words.

[1] The evidence presented by Williams (1976) is slightly different from many other studies on synesthetic metaphors. Whereas most other research looks at corpus frequencies, Williams looked at dictionary data and word etymologies.

[2] Once asymmetry (as opposed to unidirectionality) is at stake, one needs to consider whether a piece of asymmetry stems from a preference for being hierarchy-consistent or a dispreference for being hierarchy-inconsistent. A preference for A > B or a dispreference for B > A could lead to the same observed pattern, even though they would reflect different underlying mechanisms.

Chapter 9. Explaining the hierarchy of the senses

9.1. A multicausal approach

The last chapter introduced the hierarchy of the senses. This chapter looks at factors that may explain the observed asymmetries between sensory words. An overview of explanatory accounts is shown in Table 2. These accounts are not mutually exclusive, although a given study most often focuses on only one account. Research on synesthetic metaphors generally does not emphasize the possibility that metaphorical asymmetries may arise from multiple conjoining factors (for notable exceptions, see Ronga et al., 20120; Strik Lievers, 2015; Tsur, 2008, 2012).

Table 2

Summary of different explanatory accounts of the hierarchy

Account	**Explanation**
Perceptual	Language-internal asymmetries arise from perceptual asymmetries
Lexical	Asymmetries arise from differential ineffability
Evaluative	Evaluative adjectives are more likely sources
Gradability	Gradable adjectives are more likely sources
Iconic	Iconic adjectives are less likely sources
Idiosyncratic	Asymmetries arise from individual preferences and/or cultural fashions

One should distinguish between global and local accounts of explaining the hierarchy, which is a distinction that runs across the different explanatory accounts. Global accounts try to explain the entire hierarchy with one principle or a small set of principles that span across the different sensory modalities. Local accounts focus on specific crossmodal mappings—for example, providing a separate explanation for the asymmetry between taste and smell than for the asymmetry between touch and taste.

Chapter 9.2 will review the different approaches, and Chapter 9.3 will evaluate these different approaches. In Chapter 9.4, I will draw some interim conclusions and argue for the idea that the hierarchy may have multicausal origins rather than monocausal ones.

9.2. Overview of explanatory accounts

9.2.1. Perceptual accounts

By far the most common explanatory accounts are perception-based. Perhaps the most dominant account that links metaphorical asymmetries to perceptual asymmetries comes from Shen and his colleagues, who have proposed what they have variously labelled the "directionality principle" (Shen & Gil, 2007), the "cognitive principle" (Shen & Aisenman, 2008), and the "conceptual preference principle" (Shen & Gadir, 2009). Shen and Cohen (1998) talk about a "cognitive constraint," according to which a "*mapping from a more accessible concept onto a less accessible one is more natural than its inverse*" (p. 123, emphasis in original). Across Shen's body of work, this notion of accessibility has also been referenced to as conceptual preference, concreteness, or salience (Shen, 1996, 1997, 2008; Shen & Aisenman, 2008; Shen & Cohen, 1998; Shen & Gadir, 2009; Shen & Gil, 2007).

Shen and Aisenman (2008, pp. 111–113) characterize their notion of accessibility as follows: Touching, tasting, or smelling an object entails being close to it. [1] Vision and audition, on the other hand, are relatively more distal (i.e., humans can use them to experience objects from very far away).[2] This makes touch and taste in particular very "direct," because these two modalities involve immediate contact with an object. In addition to the criterion of distance, the directionality principle is thought to be grounded in phenomenological differences in the felt experience of the different sensory modalities. Experiencing something through vision and hearing is argued to be more object-based; that is, the object external to one's body is understood by the experiencer as the cause of his or her sensation (see also Shen, 2008, p. 302). Touch, taste, and smell, on the other hand, are argued to be relatively more subjective. These three senses are phenomenologically experienced as physiological sensations that are more directly connected to one's body rather than to external objects. These ideas are echoed by Pasnau (1999, p. 314), who says that "only sight and hearing perceive things as being out in the world, at a distance from the body."

A defining feature of the directionality principle is that accessibility (or its sister notions of concreteness and salience) is understood to allow separating the sensory world into "lower" and "higher" senses. As such, the hierarchy is monolithic, in that one principle (accessibility) is thought to account for the entire hierarchy. Rather than focusing on specific pairs of senses, Shen and colleagues seek to ground the hierarchy of the senses in a one-size-fits-all principle. This makes the directionality principle a global explanatory account of the hierarchy, not a local one.

Along with Shen and colleagues, other researchers have also proposed explanations that are based on perceptual asymmetries. Williams (1976, pp. 464–465) talks about more or less "differentiated" sensory modalities. Similarly, Tsur (2008) talks about the need for poetry to achieve "undifferentiatedness" (p. 286). He argues that because visual shapes are stable and differentiated, they "resist smooth synaesthesia" (p. 288).

The perceptual accounts discussed so far attempt to account for all or many different asymmetries in one swoop. A different approach is to look at the specifics of each pair of crossmodal connections (such as taste/smell or touch/taste) and see whether there is any evidence for asymmetry in perception. These are local accounts that explain only specific subparts of the hierarchy. Lehrer (1978), for example, explains the asymmetry between touch and taste as follows:

> Since we have touch receptors in our mouth, it is easy enough to understand how touch sensations could be transferred to taste. On the other hand, since we can touch things with receptors all over our skin and do not have taste receptors all over, it makes somewhat less sense to talk about something feeling *sweet* or *sour*. (p. 119)

Shibuya et al. (2007, p. 217) think that touch-to-sight asymmetry comes from the fact that “tactile sense often involves the use of the visual sense,” but not vice versa. They say that “looking at a chair gives one some visual information about the object, but it does not provide any tactile information.” To explain touch-to-sound asymmetry, Ronga et al. (2012, pp. 141–142) allude to the fact that there are asymmetries between touch and audition in perception, with touch affecting auditory perception more than the reverse (e.g., Caclin, Soto-Faraco, Kingstone, & Spence, 2002; Soto-Faraco, Spence, & Kingstone, 2004; see also Schürmann et al., 2004).

Perceptual explanations are by far the most dominant in research on synesthetic metaphors. Very few researchers even consider that the causality between language and perception may be reversed; perhaps language creates rather than reflects crossdomain mappings (see Levinson & Majid, 2014, p. 415).

There is some evidence at least that metaphorical language may influence nonlinguistic thought (Dolscheid et al., 2013), so the idea that the causal chain goes the other way—from language to perception—should at least be entertained.

9.2.2. Lexical composition and ineffability

Ullmann (1959, p. 283) considered the possibility that at least part of the observed metaphorical asymmetries could be explained through lexical differentiation (i.e., the fact that there are fewer lexical distinctions for some sensory modalities; see Chapter 4). To explain Ullmann's reasoning, it is useful to consider the connection between vision and audition. Ullmann (1959) observed in his data that "the acoustic field emerges as the main recipient" in crossmodal metaphors (p. 283). He also noted that more visual terms are used to talk about auditory concepts (e.g., *bright sound*, *pale sound*, *dark voice*) rather than the other way around (e.g., *loud color*). His explanation of this fact was as follows:

> Visual terminology is incomparably richer than its auditional counterpart, and has also far more similes and images at its command. Of the two sensory domains at the top end of the scale, sound stands more in need of external support than light, form, or colour; hence the greater frequency of the intrusion of outside elements into the description of acoustic phenomena. (Ullmann, 1959, p. 283)

Similarly, Engstrom (1946) says that "vocabulary limitations often enforce the use of intersensal terms when there is need for discriminating between impressions on the same sensory level" (p. 10). He discusses the apparent lack of differentiation in the color vocabulary of Ancient Greeks, arguing that this in part explains why they used fewer metaphors based on colors (p. 18). The notion that the use of perception-related metaphors is motivated by the richness (or poverty) of vocabulary within a particular semantic domain is held by many researchers (see also Abraham, 1987; Holz, 2007; Ronga et al., 2012; Strik Lievers, 2015; Velasco-Sacristán & Fuertes-Olivera, 2006). Metaphor is generally seen as one of the primary ways to extend meanings in order to widen the range of concepts that can be expressed (Dirven, 1985; Ortony, 1975; cf. Lehrer, 2009, p. 19). Researchers have argued that pain, which is a "subjective and poorly delineated experience that is difficult to express satisfactorily in language" (Semino, 2010, p. 206), may need metaphors to be conveyable (Semino, 2010; Schott, 2004). The idea that metaphor can compensate for limited vocabularies is perhaps most clearly stated in Fainsilber and Ortony's (1987) inexpressibility hypothesis, which "proposes that metaphors may enable the communication of that which cannot be readily expressed using literal language" (p. 240).

It is not only the number of source terms that needs to be considered, but also the frequency with which speakers feel the need to talk about the target domain, which harkens back to the discussion of communicative need in Chapter 4. Abraham (1987, p. 162) uses the German word "Differenzierungsbedürfnis" ('need to differentiate') to explain why people want to use sensory words crossmodally. Strik Lievers (2017, p. 92) says that "the dominant role of sight and hearing in human perception...may help explain why in many languages these two modalities are the most frequent targets of synaesthetic transfers." Tsur (2007, p. 34) says "the richer the sensory domain, the more it 'borrows'; the poorer the domain, the more it 'lends'." If indeed sight and sound are more important to humans, then we expect that they will be more frequently verbalized, which is what the evidence suggests (Chapter 15). Given frequent reference to visual and also sometimes auditory concepts, and given a basic linguistic need to be expressive and occasionally even extravagant in one's descriptions (see Haspelmath, 1999), we expect that speakers would have to innovate visual and auditory vocabulary more frequently. This would create a drive for borrowing lexical material from the other senses to allow continued expressivity (compare Engstrom, 1946, p. 10).

9.2.3. Evaluation

As was already mentioned in Chapter 7, many researchers have discussed a potential role of affect and evaluation in the crossmodal use of sensory words (Barcelona, 2003, 2008; Lehrer, 1978; Marks, 1978, pp. 216–218; Osgood, 1981; Popova, 2005; Sakamoto & Utsumi, 2014; Tsur, 2008, Chapter 10). Tsur (2012, p. 230) discusses that for the expression *loud perfume*, the negative meaning of obtrusiveness is more salient than the sensory impression of loudness (cf. Barcelona, 2003, 2008; Tsur, 2008). As was argued in Chapter 7, expressions that combine seemingly dissimilar sensory modalities, such as *sweet melody*, may be motivated by the evaluative nature of *sweet*. Perhaps the high emotionality of taste words (see Chapter 16) explains why taste vocabulary is often used to talk about the other senses.

9.2.4. Gradability

Scalar gradability characterizes concepts which can be conceived of as being more or less. For example, a surface can be more or less rough. On the other hand, a color term such as *blue* is not gradable. It is possible to say that one color is *more blue* than another color, but in doing so the word *more* is used to indicate a shift toward a more prototypical blue, rather than an increase in the quantity of 'blueness' (Gärdenfors, 2014, p. 136). O'Malley (1957, p. 394) already said that "rough scalar arrangements of sense qualities" may motivate "intersense analogies" (see also Abraham 1987, p. 163).

Werning et al. (2006) conducted a naturalness rating study with German crossmodal expressions. Petersen et al. (2008) noted that color terms were less acceptable in crossmodal expressions, which they link to the lack of gradability of color words. Ronga et al. (2012) observed that shape descriptors such as *round* are also not gradable, stating that "an object may or may not be round, but it cannot be more or less round" (p. 148; see also Miller & Johnson-Laird, 1976, pp. 355–357).[3] The fact that vision is associated with many non-scalar concepts may explain why vision is positioned at the top of the hierarchy. The same goes for sound, which also appears to have many non-scalar terms, such as *squealing* and *barking*.

Popova (2005) proposed that touch words are more gradable, saying that expressions such as *soft voice* come with an "implicit degree of intensity" (p. 410). According to Popova (2005), touch has more affordance for gradability because the manual exploration of surfaces is a piecemeal affair in which information is arrived at in a sequential fashion (cf. Bartley, 1953, p. 401; Ullmann, 1959, p. 283; see also Ronga et al., 2012, p. 148).

Temperature terms are also gradable and can be arranged in a linear fashion, such as *hot*, *warm*, *lukewarm*, *cool*, and *cold* (see Koptjevskaja-Tamm, 2015; Lehrer, 1978, p. 100). Abraham (1987, p. 186) notes in passing that temperature as a perceptual domain has more graded terms than other domains. Perhaps as a result of their graded nature, temperature terms are also frequently used crossmodally, such as with the expressions *warm color*, *cool color*, *warm sound*, and *cool sound*. Correspondingly, Ullmann (1959) positions temperature (separately from touch) at the bottom of the hierarchy of the senses.

Thus, altogether there is some suggestion that those sensory modalities at the bottom of the hierarchy (particularly touch) are associated with more scalar terms than those at the top of the hierarchy, which may have an overall higher number of non-scalar terms.

9.2.5. Iconicity

Adjectives are intrinsically underspecified, requiring a head noun to gain full meaning (Diederich, 2015, Chapter 4; Paradis, 2000). Some researchers have argued that crossmodal uses of sensory words require semantically malleability (compare Abraham, 1987, p. 158; Baumgärtner, 1969, pp. 16–17). For example, the fact that the word *sweet* can be used in very different kinds of contexts (such as *sweet taste* compared to *sweet melody*) suggests that the word is semantically malleable.

It is possible that this malleability may be restricted through iconicity (Lupyan & Winter, 2018). Words such as *squealing*, *hissing*, and *booming* are very strongly tied to particular auditory impressions because they are depictive (i.e., the sound of these words is highly evocative of the actual percept). This may explain why such expressions as *squealing color*, *hissing taste*, and *beeping smell* do not appear to be felicitous. The auditory word *loud* is arguably less iconic than the words *squealing*, *hissing*, and *booming*. Correspondingly, it is easier to use this word in crossmodal expressions such as *loud perfume* and *loud color* (Barcelona, 2003; Taylor, 1995, p. 139; Tsur, 2012). The idea that spoken iconicity restricts crossmodal use goes back to Classen (1993), who wrote that "auditory terms are too echoic or suggestive of the sounds they represent to be used to characterize other sensory phenomena" (p. 55). This idea will be tested in Chapter 17.

9.2.6. Idiosyncratic explanations

Ullmann (1959, Chapter 5.2) also considered the role of "the *mental make-up*, temperament, and synesthetic leanings and experiences of the individual" (p. 286, emphasis in original), the use of stimulants (such as hallucinogenic drugs), and "contemporary *fashions*, literary and artistic models, and current modes of vision" (p. 288, emphasis in original). Tsur (2007) discusses the many ways writers and poets may violate certain expectations in what he calls "literary synesthesia" to create witty effects. However, as a stylistic device, this trope is not without its critics (Engstrom, 1946), and its use differs across literary traditions. Earlier work on the crossmodal uses of sensory words focused more strongly on highlighting differences between poets and stylistic traditions (Engstrom, 1946; Erzsébet, 1974; Ullmann, 1945, 1959; Whitney, 1952) than modern work (but see Day, 1996; Popova, 2003; Strik Lievers, 2016). Some scholars even outright rejected the idea that there are lawful principles in crossmodal language (Abraham, 1987, p. 162).

9.3. Evaluating the different explanatory accounts

Let us take stock of the various explanatory accounts discussed in this chapter. How plausible is each account? And to what extent has each account already been supported by empirical evidence?

9.3.1. Evaluating perceptual accounts

When Williams (1976) discussed parallels between language-external aspects of perception and linguistic hierarchy, he said the following about his own speculations: "It should be strongly emphasized that the following are presented only as striking parallels, to pique interest. No cause-effect relationship whatever is claimed" (p. 472).

In fact, the same reasoning applies to *all* arguments based on language-external factors, so long as no experimental link between metaphor usage and perceptual factors is directly established. A host of experimental and neurophysiological findings link different senses with each other (e.g. Spence, 2011), often in context-dependent and phenomenon-specific ways. There are myriads of differences between the senses, and there are many different experimental findings that suggest various forms of asymmetries between the senses.

Any account based on language-external perceptual facts is open to attack from an argument based on an alternative set of facts. Consider that Shibuya et al. (2007) argue for a linguistic touch-to-sight asymmetry based on the fact that in perception, vision dominates touch (e.g., Hay & Pick, 1966; Rock & Victor, 1964) wheareas Ronga et al. (2012) provide a related argument, but they apply the opposite logic. These authors cite experimental results where touch dominates sound in perception (e.g., Caclin et al., 2012) to argue for a linguistic touch-to-sound asymmetry. What is more, both researchers only address aspects of the asymmetry between perceptual modalities. The dominating sense in perception depends on the task and the particular qualities investigated. For example, it has been found that vision dominates sound with respect to spatial perception, but sound dominates vision with respect to temporal perception (Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Which of these perceptual asymmetries shall be taken to motivate linguistic asymmetries? It is impossible to answer this question without further evidence.

Perceptual accounts of the monolithic nature (global accounts) are even more problematic, particularly the above-mentioned idea put forth by Shen and colleagues that the more “accessible” senses are mapped onto the less “accessible” ones. Rakova (2003) remarks that their notion of accessibility is "suspect" (p. 113). Indeed, it appears to be at odds with most empirical work on accessibility. Psycholinguistically, this construct is generally operationalized in terms of speed of processing. Perceptual experiments show that people are quicker at processing visual and auditory information than tactile information (Spence et al., 2001; Turatto et al., 2004), and the same applies to visual and auditory words, compared to touch words (Connell & Lynott, 2010; see also Connell & Lynott, 2014). Sometimes, the notion of accessibility is operationalized in terms of word frequency (e.g., Harmon & Kapatsinski, 2017), in which case vision would seem to be the most accessible sense because it is the most frequent (see Chapter 15). Paradis and Eeg-Olofsson (2013, p. 37) and Caballero and Paradis (2015) state that the directionality principle is at odds with the crosslinguistic literature on the evidentiality hierarchy, according to which vision is the most accessible sensory modality (which explains why visual terms often form the basis of grammaticalization processes that lead to evidential markers).

In general, the notion that the sensory world can be categorized into lower and higher senses is not grounded in any neuropsychological facts. Most scientists working on the senses do not accept this division, and some have vehemently argued against it.[4] Furthermore, the notion that some senses are more or less differentiated than others—seen in such proposals as Williams (1976) and Tsur (2008, 2012)—also does not map straightforwardly onto anything in sensory science; there is no widely accepted empirical measure of perceptual differentiatedness. The vagueness of such overarching concepts is illustrated by Rakova (2013, p. 113), who says that touch might just as well be the most rather than the least differentiated sense, given that it is associated with numerous different perceptual dimensions (pressure, pain, temperature), each with its own type of physiological subsystem.

However, a deeper problem with the work of Shen and colleagues (e.g., Shen, 1997; Shen & Cohen, 1998; Shen & Gil, 2007, etc.) on the directionality principle is that linguistic evidence is treated as evidence for a particular perception-based explanatory account without actually incorporating any genuine perceptual evidence. For example, Shen and Cohen (1998) see their cognitive constraint confirmed by a series of experiments showing that participants perceived hierarchy-consistent metaphors to be more adequate than hierarchy-inconsistent metaphors, and participants also found the latter more difficult to understand. Similarly, Shen and Aisenman (2008) discuss experiments showing that hierarchy-consistent metaphors are judged to be more natural and more memorable. They see this as further evidence for the directionality principle (p. 110), although the observed effects would be consistent with many different explanatory accounts.[5] There is nothing wrong with this experimental evidence per se, which does in fact lend further support to the hierarchy as a descriptive generalization. However, nothing in these experiments actually necessitates the assumption that the asymmetries are grounded in perceptual accessibility.

Similarly, Shen and Cohen (1998) cite the exclusively linguistic evidence in Williams (1976) as "supporting the claim that lower sensory domains are more accessible" (p. 129), even though the study of Williams (1976), which only looked at lexicographical evidence, has nothing to say about accessibility as a cognitive notion. Shen and Cohen's reinterpretation of Williams (1976) deviates from his own carefully chosen words that "no cause-effect relationship whatever is claimed" (p. 472).

The linguistically and psycholinguistically observed asymmetries may well be grounded in some form of accessibility, but the evidence presented by studies such as Shen and Cohen (1998) and Shen and Aisenman (2008) alone does not allow this conclusion. Amassing more evidence for asymmetries in corpus frequencies or asymmetries in psycholinguistic processing speaks to the generality of the hierarchy as a descriptive generalization, but it does not speak to any one explanatory account, such as the alleged directionality principle. In contrast to conceptual metaphor theory, for which there is abundant nonlinguistic evidence (see, e.g., Boroditsky & Ramscar, 2002; Casasanto & Boroditsky, 2008; Landau et al., 2010), a lot of research on synesthetic metaphors has not addressed the circularity concerns that haunted early work on conceptual metaphors (cf. Gibbs, 2007; Murphy, 1996, 1997) and continues to haunt cognitive linguistics to this day (Dąbrowska, 2016a).

With some noteworthy exceptions (such as Ronga et al., 2012), research on the hierarchy of the senses does not engage in a deep fashion with experimental evidence from other fields to supports its claims. Sensory modalities are claimed to be more or less “concrete,” “differentiated,” “stable,” or “distant”—none of which are concepts that are widely accepted in psychological work on the senses. In fact, research on synesthetic metaphors incorporates cultural categories such as the five senses folk model and the idea of “higher” and “lower” senses into its explanatory accounts without questioning whether and how these cultural categories map onto sensory perception.

I am only aware of one attempt to directly link linguistic and perceptual evidence experimentally. This attempt specifically addresses the connection between sound and sight, as revealed in such expressions as *the murmur of the gray twilight* and *the sound of coming darkness* (from Edgar Allan Poe; taken from Marks, 1982b, p. 16). Marks (1982b) showed that participants reliably matched such expressions as *dark cough* and *bright sneeze* to the loudness of an auditorily presented sound stimulus or to the brightness of a visually presented light stimulus. In addition, brightness and loudness judgments were correlated with each other. This corresponds to independent perceptual evidence that people reliably match brightness to loudness in perception (Marks, 1978; Marks & Stevens, 1966; Stevens & Marks, 1965). This evidence thus links expressions such as *dark cough* and *bright sneeze* to an underlying crossmodal correspondence. This explanation is a local one, however, only accounting for audiovisual language.

In sum, there are numerous problems with perception-based accounts, including the lack of incorporation of actual perceptual studies, the lack of theoretical constraints on positing language-perception connections, and the problem of circular reasoning. Only carefully conducted experiments that link language with actual perceptual stimuli, such as conducted by Marks (1982b), can be used to support perception-based accounts.

9.3.2. Evaluating ineffability-based accounts

Tsur (2012, p. 227) is skeptical of Ullmann's original explanation that differential lexicalization may be a driving factor behind metaphorical asymmetries, calling the proposal "not very convincing" because "poverty of terminology is not the only (or even the main) reason for using metaphors in poetry" (see also Tsur, 2007, p. 33; Tsur, 2008, p. 285). First, it must be emphasized that it was never in question whether "poverty of terminology" was the only or even the main reason for asymmetries in crossmodal language use. Researchers such as Strik Lievers (2015), for example, have considered the possibility that poverty of terminology is *one of many* contributing factors.

In fact, there is already initial empirical evidence for a role of lexical differentiation in explaining metaphorical asymmetries: Strik Lievers (2015, pp. 86–88) showed for her English, Italian, and German datasets that modalities associated with a high number of adjectives were more likely sources in what she calls "synaesthesia," whereas modalities associated with a high number of nouns were more likely targets. A similar observation was made by Ronga and colleagues (2012). These authors replicated the frequently observed touch-to-sound correspondence for Italian, but they also observed that this may stem from an already-existing asymmetry in the lexicon: In their dataset, touch had more adjectives and fewer nouns than sound (Ronga et al., 2012).

It should furthermore be noted that an explanation of the hierarchy grounded in already existent imbalances in the sensory lexicon is consistent with numerous already established facts. First, Fainsilber and Ortony (1987) provide experimental evidence showing speakers use more metaphors when they describe phenomena that are otherwise difficult to describe. Second, there is evidence suggesting a greater communicative need to talk about visual concepts in English (Winter et al., 2018), which may lead to an increased frequency of borrowing from other perceptual domains. Third, there is evidence suggesting that sound in particular has few adjectives, consistent with the fact that these adjectives are infrequently used in crossmodal constructions (Strik Lievers, 2015; Strik Lievers & Winter, 2018; Ronga, 2016). On the other hand, touch is frequently found to have many adjectives, consistent with the fact that touch vocabulary is frequently borrowed. Smell, another modality that is infrequently used as a source, has also been reported to have few adjectives (as noted by Tsur, 2008, p. 288). Fourth and finally, Chapter 17 will provide new evidence for the idea that the differential ineffability of the senses plays a role in the hierarchy of the senses. I will show that differential lexicalization alone can create patterns that look almost exactly like the patterns that are commonly used to support the hierarchy of the senses.[6]

So, contrary to Tsur, we may conclude that even highly creative and expressive poets such as Byron are constrained by the limits of their language. When poets generate those linguistic expressions that literature scholars call "synesthesia," they have to make do with what the language has to offer to them. In fact, given the discussion of ineffability in Chapter 4, it is almost a logical necessity that Byron and other poets, just like any other language user, will experience some constraints when expressing sensory meaning, which then necessitates finding "poetic" means around these constraints.

9.3.3. Evaluating evaluation-based accounts

Several pieces of data support an evaluation-based account. The evidence presented Winter (2016) and Chapter 16 shows that taste and smell words are overall more emotional compared to descriptors from the other modalities. For vision in particular, there is a larger proportion of neutral words (e.g., *yellow*). Given that evaluation may be a driving factor in crossmodal language, as was argued in Chapter 7, this observation fits the fact that taste and smell are at the bottom of the hierarchy as likely sources.

In addition, an evaluation-based account is broadly consistent with the increasing recognition that metaphors in general may be more emotionally engaging (see Citron & Goldberg, 2014; Citron, Güsten, Michaelis, & Goldberg, 2016; but see Pomp et al., 2018), and that crossmodal language use also has emotional effects (Sakamoto & Utsumi, 2014). Edwards and Clevenger (1990), for example, provide experimental evidence showing that when speakers have stronger emotional engagement with a topic, they are more likely going to use metaphors to describe it. Chapter 17 will provide additional evidence for a direct link between the evaluative meaning of words and their propensity to be used in crossmodal expressions.

9.3.4. Evaluating gradability-based accounts

Although several researchers have informally observed that gradable adjectives are more likely used crossmodally (Petersen et al., 2008; Popova, 2005), I am not aware of any quantitative evidence that supports this position. Gradability-based accounts are, however, quite plausible given that in perception (independent of language) people find it easy to match perceptions from different modalities based on graded scales of intensity and magnitude (Marks, 1978; Stevens, 1975; see also Winter et al., 2015b).

9.3.5. Evaluating iconicity-based accounts

There is a clear correspondence between the fact that sound concepts are unlikely sources in crossmodal expressions and the well-established fact that sound concepts are more iconic (Dingemanse, 2012; Perlman et al., 2018; Winter et al., 2017; see also Chapter 15); however, these two facts have so far not been directly related. This will be achieved in Chapter 17.

The iconicity-based account is furthermore consistent with ideas explored under the banner of the "double mapping constraint" in sign language linguistics (Emmorey, 2014; Meir, 2010; see also Lupyan & Winter, 2018). It has been observed that the particular iconic mapping a sign selects may block certain metaphorical expressions. For example, Israeli Sign Language and American Sign Language are reported not to have the metaphorical expression *Time flies* because in both languages, the sign for the concept 'flying' involves mimicking bird fly. Thus, the iconicity of this sign depicts only a particular type of flying, and this type of flying is incompatible with the idea of time flying. Consistent with the general idea that iconicity restricts semantic extension (e.g., via metaphor), Lupyan and Winter (2018) show that iconicity ratings are negatively correlated with measures of contextual diversity (i.e., how many different types of contexts a word occurs in): Relatively more iconic words occur in less distinct text types.

9.3.6. Evaluating idiosyncratic explanations

Canobbio (2004; discussed in Strik Lievers, 2015, p. 73) examines the poetic works of Corrado Govoni, noting that in this poet's work, most crossmodal uses of sensory words disobey the hierarchy. This is a demonstration of a particular writer whose work is not in line with the hierarchy of the senses, showing that individual and stylistic factors are important. However, while individually and stylistically contingent factors are clearly relevant, they will not be considered here since they do not allow for strong generalizations about usage patterns that are found in the general population.

9.4. The multivariate road ahead

In a paper on metaphors and perception, Marks (1996) writes the following:

> Even if some perceptual metaphors end up being mediated linguistically, their origins appear to be wholly in perception itself, starting within perceptual processes before being overlaid and dominated by linguistic ones. (p. 59)

The problem with this perspective is that it makes the claim about the relative strength of two factors, perception on the one hand, and language on the other, without any direct empirical evidence about the relative import of these two factors. There simply are no empirical assessments that actually test the strength of one explanatory account against the strength of another. In fact, since most researchers have focused on perceptual accounts, there currently is no quantitative evidence for many of the hypothesized linguistic factors, such as iconicity and adjective scalarity. Given this scarcity of empirical evidence, it is perhaps prudent, for the time being, not to deem any one explanatory account dominant. However, Chapter 17 will provide novel empirical evidence establishing that at least three different factors conjointly play a role. These factors are: the composition of the sensory vocabulary, emotional valence, and iconicity.

This is not to say perception-based explanations should not be investigated, let alone ruled out. In fact, if the Embodied Lexicon Hypothesis (Chapter 5) is true, we would expect to find correspondences between perceptual asymmetries and asymmetries in crossmodal metaphor use. However, given the many different crossmodal connections involved in the hierarchy, it is important not to jump to conclusions and to await further testing. Moreover, language-internal explanations should be explored alongside perception-based explanations.

It should be noted that although Ullmann's original work (1945, 1959) is generally cited as evidence for a monolithic hierarchy, his writing actually reflects a willingness to consider a multitude of explanatory factors that has since been lost in much subsequent work. Few researchers (including Ronga et al., 2012; Strik Lievers, 2015; Tsur, 2008, 2012) consider the possibility that there may be multiple factors involved.

Mitchell (2004, p. 81) states that "there will never be a single account that can do all the work of describing and explaining complex phenomena." Similarly, Markman (2008, p. 247) states that "it is often easier to envision a complete architecture for cognition emerging from the merger of approaches." Thus, rather than focusing on a monolithic hierarchy, a more productive research strategy may be to ask the question: What multitude of factors cause asymmetries in the crossmodal uses of sensory words?

Even given the present lack of evidence, it is already apparent that some aspects of the hierarchy may best be explained by a combination of explanatory approaches. For example, many of the different factors stated in this chapter conspire against sound being used as a source. Most sound concepts appear to be non-scalar (e.g., *squealing*) as well as highly iconic (e.g., *beeping*), both of which may bias against their crossmodal use. In addition, sound concepts are also not particularly differentiated, as Chapter 12 will show. These facts are consistent with a multi-causal view of the hierarchy of the senses, but a formal test of these different factors still awaits.

---

[1] There are slight differences in how the notion of accessibility has been characterized across the body of Shen's work. Shen (1997) also included the criterion of whether there is a dedicated sensory organ associated with a particular sensory modality. In Shen's (1997) notion of accessibility, part of the evidence for touch being "low" is that it "does not use a special organ" (p. 54). This, however, is wrong since touch does have a special organ, the skin (cf. Carlson, 2010; Møller, 2012). This point was already made by Rakova (2003), who said the following in response to Shen (1997): "The fact that the sense of touch is realized by cutaneous receptors which cover the entire body surface does not entail that there is no special organ for touch" (p. 113).

[2] Already Aristotle criticized the proximal versus distal criterion, observing that all sensory perception involves contact (Aristotélēs, *De Anima*, 423b1, in Macpherson, 2011):

> The problem, then, is: does the perception of all objects of sense take place in the same way, or does it not, e.g. taste and touch requiring contact (as they are commonly thought to do), while all other senses perceive over a distance? The distinction is unsound; we perceive what is hard or soft, as well as the objects of hearing, sight, and smell, through a

medium, only that the latter are perceived over a great distance than the former; that is why the facts escape our notice. (p. 56)

The idea that vision is more distal than the other senses is not universally agreed upon. Dominy and colleagues (Dominy, Lucas, Osorio, & Yamashita, 2001) describe how for primates living in arboreal environments, both hearing and smelling are more far-ranging than sight. On the other hand, Speed and Majid (2017) find mental simulation effects consistent with the idea that olfaction is mentally represented as close to the body.

[3] Shape descriptors are rated to be primarily visual by the native speakers in the Lynott and Connell (2009) study.

[4] Shen and Cohen (1998) state that a linear scale of the senses (from low to high) is "commonly assumed" (p. 125), citing authors such as Ullmann (1957), Tsur (1992), and Williams (1976). However, all these authors are language researchers. Many researchers outside of linguistics actually do *not* assume a linear scale of the senses, including anthropologists (Classen, 1993; Howes, 1991).

[5] In fact, the very existence of hierarchy-consistent asymmetries in corpus data poses problems for the interpretation of these experiments. The fact that participants perceive hierarchy-consistent metaphors to be more natural may stem from the increased frequency of these expressions.

[6] A further hypothesis to be maintained is that the differential lexicalization of the senses has downstream behavioral effects. In particular, it is quite plausible that because some modalities have more adjectives associated with them, language users become accustomed to borrowing terminology from these modalities.

Part II.
A case study of sensory adjectives

Chapter 10. Methodological foundations

10.1. Theory and method

Methods form the bedrock of science. Werner Heisenberg (1958, p. 58) famously said: "We have to remember that what we observe is not nature in itself but nature exposed to our method of questioning." Method and theory depend on one another, and "no method...is neutral with regard to theory" (Hunston, 2000, p. 250).

Methods also form the bedrock of interdisciplinary work. Researchers from different fields can only collaborate with each other if they achieve their respective standards of evidence. Sharing a common methodological foundation, such as established statistical techniques that are employed across the sciences, permits the building of bridges between scientists from different backgrounds and different research traditions. The connection between methods and interdisciplinary work is especially important for sensory linguistics because of its intrinsically interdisciplinary nature.

Among other things, this book sets out to introduce new methods for studying natural language data. Because the history of the language sciences has repeatedly shown that work can easily be disregarded because of differences in method, this entire chapter is devoted to outlining the book's methodological commitments. The empirical chapters that follow will adhere to these commitments.

Perhaps Goldinger et al. (2016) exemplifies how important work within the language sciences can be ignored because of methodological differences. These authors argue that embodied cognition is irrelevant for the cognitive sciences based on a review of the experimental literature on cognitive processing in which they ignore all of cognitive linguistics. Most classic research in cognitive linguistics (e.g., Lakoff & Johnson, 1980; Langacker, 1987, 2008) relies exclusively on the intuition of the researcher, something that Dąbrowska (2016a) has called a "deadly sin" of cognitive linguistics. Perhaps as a result of this, cognitive psychologists (such as Goldinger et al., 2016) do not pay enough attention to the body of work in this domain (for discussion, see Gibbs, 2007).

This chapter proceeds as follows. To start this discussion of methods, Chapter 10.2 introduces the three commitments of cognitive linguistics, which also form guiding principles for sensory linguistics. Chapter 10.3 introduces a new commitment, the Reproducibility Commitment, which will be illustrated with two problematic cases in Chapter 10.4. One solution to the problematic cases, and one way of furthering reproducibility in the language sciences, is to incorporate already existing data from rating studies ("norms") into linguistic analyses (Chapter 10.5). How such norms fit into general concerns for theory-building is discussed in Chapter 10.6. Finally, the statistical software used throughout Part II of the book will be described in Chapter 10.7.

10.2. Cognitive linguistic commitments

One way to characterize cognitive linguistics is by detailing the methodological–theoretical commitments it adheres to. Following Lakoff (1990) and Lakoff and Johnson (1999), three different principles or "commitments" form the bedrock of any methodological considerations in this book: The Cognitive Reality Commitment, The Convergent Evidence Commitment, and The Generalization and Comprehensiveness Commitment. These commitments are generally taken to be core commitments for cognitive linguistics, guiding principles for both theorizing and data collection. However, in many ways they form desiderata for the language sciences in general. The following descriptions of the commitments are taken straight from Lakoff and Johnson (1999, pp. 79–80):

> *The Cognitive Reality Commitment:* "An adequate theory of concepts and reason must provide an account of mind that is cognitively and neurally realistic."
>
> *The Convergent Evidence Commitment:* "An adequate theory of concepts and reason must be committed to the search for converging evidence from as many sources as possible."
>
> *The Generalization and Comprehensiveness Commitment:* "An adequate theory must provide empirical generalizations over the widest possible range of phenomena."

The Cognitive Reality Commitment requires that theories in the language sciences are consistent with relevant existing work, such as work on psycholinguistic processing and cognitive processing more generally. In essence, the Cognitive Reality Commitment expresses a very fundamental property of all of science, which is that "all relevant knowledge should be brought to bear on interesting problems" (Dienes, 2008, p. 7). Linguistic theories cannot and should not be sitting in a void, aloof from the rest of science.

Sensory linguistics needs to pay special attention to the Cognitive Reality Commitment; it relies heavily on insights from such fields as psychophysics, cognitive psychology, neurophysiology, and anthropology. Unfortunately, researchers working in the domain of sensory linguistics often only pay lip service to results from other fields, or they only incorporate extralinguistic evidence sparingly into their theories (for notable exceptions, see Lehrer, 2009; Levinson & Majid, 2014; Ronga, 2016; Ronga et al., 2012). Often, when linguists investigate sensory language, the perceptual science does not go beyond common knowledge about the senses, and often models that are not based on science (such as the five senses folk model) are tacitly incorporated into linguistic theories as if they were established facts.

Dąbrowska (2016a) criticizes cognitive linguists for not taking the Cognitive Reality Commitment more seriously. According to her, the relation of cognitive linguists toward the Cognitive Reality Commitment is best described as "believing but not practicing" (p. 482). She calls for cognitive linguists to engage more deeply with more current strands of psychological and neuroscientific evidence. This call is taken seriously here by paying special attention to the nonlinguistic literature on sensory modalities.

The Convergent Evidence Commitment is another foundational commitment of cognitive linguistics. Gries, Hampe, and Schönefeld (2005) characterize it as follows: "Hypotheses and constructs should be backed up by converging evidence from multiple sources" (p. 636). This is important because each method comes along with its own set of constraints and assumptions. By using several different approaches, it is possible to show that a specific set of findings is not just due to the particularities of one particular method. As stated by Dąbrowska (2016b), "when it comes to understanding something as complex as human language, it will be most productive to use every method that is available" (p. 57). The seminal volume *Methods in Cognitive Linguistics* (Gonzalez-Marquez, Mittelberg, Coulson, & Spivey, 2007) showcases how cognitive linguistics can use a wide array of methods to support its theories, including corpus analysis, psycholinguistic experiments, gesture research, and neuroscience.

The idea of converging evidence not only means that results should be triangulated via different methods, but also via different datasets. For example, Winter (2016) analyzes the evaluative functions of sensory words and observes that taste and smell words are overall more emotional (see also Chapter 16).

Crucially, emotional meaning was operationalized in three different ways, using emotional valence data collecting from a human rating study (Warriner et al., 2013), as well as valence data that was generated using dictionary-based or corpus-based methods. Statistical analyses of the three different datasets converged on the same result.

The final commitment put forth by Lakoff and Johnson's (1999) is The Generalization and Comprehensiveness Commitment. As stated by Markman (2008), "theories that cover a broader range of data are to be preferred to those that cover a narrower range of data" (p. 247). The Generalization and Comprehensiveness Commitment is related to, but slightly different from, the Convergent Evidence Commitment, which is about the same theory supported by multiple different methods. The Generalization and Comprehensiveness Commitment states that a given theory should be able to account for as many facts as possible, and preferably for as many different classes of facts (see Thagard, 1978).

The next section will introduce an additional commitment, the Reproducibility Commitment, which is especially important for the work undertaken in Part II of this book.

10.3. The Reproducibility Commitment

The Reproducibility Commitment is formulated here as follows:

> *The Reproducibility Commitment:* "An adequate theory of linguistics needs to be supported by evidence that can be reproduced by other linguists who did not conduct the original study."

In many different areas of science, there is currently a great push toward increasing reproducibility (Gentleman & Lang, 2007; Mesirov, 2010; Munafò et al., 2017; Peng, 2011), which needs to be distinguished from replication. The latter refers to the process of conducting a new study following the procedures laid out by a previous study. On the other hand, reproducibility means that *the entire process* of conducted research can be reproduced, including study design, data collection, data processing, and statistical analysis. At its minimum, reproducibility requires that a different researcher, given the same data, can reach the same conclusions. This requires making both the data and the analysis code publicly available. Such reproducibility is often not adhered to in the language sciences. For example, Winter (2011) found that statistical reporting in phonetics was so insufficient that it was impossible to even verify whether some of the most basic statistical assumptions (such as the independence assumption) were violated or not. Often p-values were reported without even stating which test was used, making it impossible for other researchers to reproduce the statistical analyses.

There is a myriad of decisions an analyst must make in a given data analysis—what Gelman and Loken (2013) talk about as the "garden of forking paths" and Simmons, Nelson, and Simonsohn (2011) describe as "researcher degrees of freedom." A particularly striking demonstration of the effects of researcher degrees of freedom was conducted by Silberzahn et al. (2017). These authors gave the same dataset to 29 analysis teams. All were given the same research question. Twenty teams found a statistically reliable support for the research hypothesis in question; 9 found no such support. In addition, those teams that did find statistically reliable effects reported quite varying effect sizes. Because statistical analysis is a subjective process (McElreath, 2016, p. 95) that involves theoretically guided analysis decisions that may differ between researchers, there is no “best analysis” in an objective sense. Consequently, researcher degrees of freedom are unavoidable. Rather than trying to eliminate them, they should be laid open.

It is important that reproducible quantitative research exists alongside qualitative approaches to studying language, and the two are not mutually exclusive. The Reproducibility Commitment says that ultimately, as few claims as possible should be based *only* on qualitative analysis or *only* on single judgments made by a single human annotator, particularly if the annotator is himself or herself a language scientist. The words “ultimately” and “as few ... as possible” are important hedges in the above statement, intended to allow for wiggle room. The demand is not, of course, that everybody has to do quantitative work and follow a reproducible research workflow. As highlighted by Fiedler (2017):

> Science is a pluralistic endeavor that should not be forced into the corset of one specific format. If science is to flourish and to achieve progress, there must be room for competing theories, methods, and different conceptions of what science is about. (p. 57)

However, at some point, phenomena that have been sufficiently explored via qualitative analyses should be supplemented with more objective and rigorous approaches that are more easily shared with researchers within the language sciences and across the disciplines.

10.4. Reproducibility: Two examples

In this section, I will give two specific examples to illustrate the issue of reproducibility in the language sciences.

10.4.1. Synesthetic metaphors

As was discussed in Chapter 8, researchers studying crossmodal language use have proposed that there is a hierarchy of the senses governing which senses can be used to talk about which other senses; for example, *sweet fragrance* is a more appropriate expression than *fragrant sweetness* (Shen & Gil, 2007). I will use Stephen Ullmann's original data to illustrate some problems with reproducibility in this field of research. His work is very old (dating back to the 1940's and 1950's), so it is perhaps natural that given the methodological standards of the time, reproducibility in the modern sense is not assured. The purpose of this discussion is not to disqualify the truly visionary work of Ullmann, a lot of which has been replicated in more rigorous studies (e.g., Ronga et al., 2012; Strik Lievers, 2015), but to highlight certain methodological principles that still affect research on crossmodal language use to this day. Ullmann's inspirational work can be used to reflect on possible avenues for methodological improvements, which will be followed through in Chapter 17.

Table 3 is taken from Ullmann (1945, p. 814) and lists how often a sensory term from a given modality is used to modify a sensory term from another modality in the writings of the poet Lord Byron. Source modalities are represented as rows (i.e., they are the sensory descriptors used to modify a particular sensory target). The target modalities are represented as columns (i.e., they are the perceptual impressions being described). An expression such as *to sweeten the sounds* (Ullmann, 1945, p. 814) would add one data point to the cell that is in the taste row and the sound column.

Table 3

Crossmodal uses of sensory adjectives by Lord Byron

	Touch	Heat	Taste	Smell	Sound	Sight	Total
Touch	(-)	8	3	3	76	31	121
Heat	2	(-)	2	-	11	9	24
Taste	1	-	(-)	1	7	8	17
Smell	-	-	-	(-)	3	2	5
Sound	-	-	-	-	(-)	11	11
Sight	5	3	-	1	21	(-)	30
Total	8	11	5	5	118	61	208

All cells in the upper-right triangle of this crosstabulation are counted as being consistent with the hierarchy of the senses. All cells in the lower-right triangle are counted as inconsistent (see Chapter 17). Thus, Table 3 suggests that this dataset from Lord Byron supports the hierarchy of the senses (but see Chapter 17).

Let us now turn to issues regarding the evidence for Ullmann's original proposal—especially since some of these issues still have not been resolved in more modern research on crossmodal language. The first issue with Table 3 is the question as to where the counts come from—what is the source of data? We know that the data comes from Lord Byron, but we do not know which specific examples are treated as "metaphorical" or not. This is particularly problematic because Ullmann (1945) gives himself a lot of freedom with respect to what is or is not counted as metaphorical. He says that he only analyzed "examples of clear synaesthesia" that are "vividly felt as such," excluding "stale epithets" such as *sweet sound* and *soft color* (p. 814). We cannot know what Ullmann did or did not take to be "clear" or "stale" cases of synesthetic metaphor. Moreover, given that he searched Byron's texts by hand, Ullmann could have easily missed certain occurrences or picked examples in a way that was biased by his theory.

We also do not know whether the table represents type or token frequencies; that is, if the word *sweet* was used multiple times to modify the word *melody*, were these counted as separate instances? By not distinguishing between type and token counts in data summaries such as Table 3, it is impossible to know whether the results perhaps overly depend on a few highly conventionalized lexical items (see discussion in Ronga, 2016). For instance, could it be that all counts in the taste-to-sound column involve the adjective *sweet*, such as *sweet melody*, *sweet music* and *sweet voice*? If that is the case, then this would considerably reduce our confidence in there being a general asymmetry between taste and sound.[1]

A final problem with the data presented in Table 3 has to do with how the analyst associates particular words with specific sensory modalities. For example, Day (1996) lists the adjective–noun pair *heavy explosion* as a touch-to-sound mapping—even though the adjective *heavy* is a general magnitude term and even though *explosion* is a noun that denotes something that can be seen, heard, felt, and smelled. Perhaps it was determined by Day (1996) that *heavy explosion* was a touch-to-sound mapping in the particular use context, but this is impossible to tell without actually being able to look at the linguistic contexts from which this example was picked. Similarly, consider the fact that Sakamoto and Utsumi (2014, p. 2) classify the adjective *open* as not being perceptual at all, even though the common sensible 'openness' denoted by this adjective can clearly be perceived through vision and touch and would probably be treated as a dimension word in William's (1976) study. Unfortunately, the fact that the data behind Table 3 is not accessible means that we do not know how cases such as *wide*, *narrow*, and *large*, or highly multisensory words such as *harsh*, have been dealt with. Because different researchers may classify sensory words differently (as discussed in Ronga, 2016), they may approach the same dataset, such as Lord Byron's writings, and come to different conclusions. Chapter 11 will introduce a wordlist of modality norms (collected via human ratings) that alleviates these concerns.

Again, it should be stated that it is easy to criticize old work that was truly revolutionary for its time but might not be able to live up to modern standards of linguistic methodology, especially given the fact that reproducibility is a relatively recent topic in science (Gentleman & Lang, 2007; Mesirov, 2010; Munafò et al., 2017; Peng, 2011). However, it must be acknowledged that given these methodological concerns, we are now somewhat less certain as to whether a hierarchy of the senses is, in fact, supported by the data. Luckily, some researchers have found converging evidence for the hierarchy using more rigorous methods. Ronga et al. (2012), Strik Lievers (2015), and Ronga (2016) start with a list of sensory terms that was assembled before approaching the corpus. This makes it explicit which words have been classified as belonging to which sense, and the existence of a pre-specified word list makes it possible for a different researcher to approach the same dataset and reach the same conclusions. Moreover, assembling a list of sensory words ahead of time reduces theoretical bias at the analysis stage.

10.4.2. Semantic prosody

Another problematic case for reproducibility involves the notion of "semantic prosody," which will be the topic of Chapter 16. This term is associated with the British tradition in corpus linguistics and was first proposed by Louw (1993), following Sinclair (1991). The idea of semantic prosody is that a word is imbued with meaning by its collocates, the frequent contexts in which it occurs (for different interpretations, however, see Hunston, 2007; Morley & Partington, 2009; Whitsitt, 2005; see also discussion in Stewart, 2010). A classic example of this phenomenon is the phrasal verb *to set in*, which often collocates with bad things in naturally occurring language, such as the words *rot* and *decay* (Sinclair, 1991). The verb *to cause* is similarly said to have negative semantic prosody by virtue of collocating with overarchingly negative words such as *war*, *damage*, *destruction*, and *chaos* (Stubbs, 2001, p. 65). A subtler case of semantic prosody is discussed in Louw (1993), who notes that in a large set of texts, the plural noun *days* appears to connote nostalgia, a longing for the past, as in *his babying days are over*, or *The big beer drinking days are gone* (p. 36).

Semantic prosody as a phenomenon has been very important for corpus linguistics because it led to a number of counterintuitive results that could only be revealed by looking at large quantities of natural language data. The fact that *cause*, for example, has negative semantic prosody is not immediately intuited when the word is presented in isolation. Only when looking at many linguistic contexts does the word reveal its semantic prosody. Thus, cases such as *cause*, *set in*, and *days* seemed perfectly suited to demonstrate the utility of corpus linguistics. For the case of *set in*, Whitsitt (2005, p. 287) discusses the surprise factor that was involved in the "discovery" of its semantic prosody as follows: "A phrasal verb, whose meaning and use was apparently familiar to all, was revealing things about itself that had not been known before."

What methods can be used to establish semantic prosody? The dominant approach in corpus linguistics, including studies on semantic prosody, is to look at concordance lines—that is, lines of text showing the head word (such as *set in*) with its surrounding contexts (see Huston, 2002, Chapter 3). This approach is seemingly more objective than what is done in other fields, such as cognitive linguistics and Chomskyan/generative linguistics, where researchers often investigate made-up examples that do not come from naturally occurring data. However, Hunston (2000, p. 65) clearly stresses that "concordance lines present information; they do not interpret it. Interpretation requires the insight and intuition of the observer." This is because, "strictly speaking, a corpus by itself can do nothing at all, being nothing other than a store of used language" (Hunston, 2000, p. 3). There always is an additional step of interpretation, which is usually done by the corpus linguist herself, as also stressed by Grondelaers, Geeraerts, and Speelman (2007):

> ...corpus research is neither automatic, nor necessarily free from the hermeneutic, interpretative features that are typical of a non-objectivist methodology... Corpus research, in fact, neither denies nor ignores the necessity of interpretations, but it takes on a helix-like structure of gradual refinement of interpretations through a repeated confrontation with empirical data. (p. 150)

The manual annotation of concordance lines is particularly problematic in the case of semantic prosody. Whether something is good or bad is intrinsically subjective; emotional meaning is notoriously difficult to pin down. Bednarek (2008, p. 122) notes that it is "difficult to establish objectively" whether a given lexical item is positive or negative. Stewart (2010, p. 91) states that "classifying co-occurrences as favourable or unfavourable is anything but straightforward, in part because what is one analyst's meat is another analyst's poison."

Because of this inherent subjectivity, concordance-based research needs to be completed with work that does not rely on manual annotation. In the case of semantic prosody, a concordance-free corpus linguistics could use established emotional valence datasets. There are by now dozens of datasets that quantify the extent to which a word is positive or negative (Baccianella, Esuli, & Sebastiani, 2010; Esuli & Sebastiani, 2006; Kiritchenko, Zhu, & Mohammad, 2014; Liu, 2012, Chapter 6; Mohammad & Kiritchenko, 2015; Pang & Lee, 2008, Chapter 7; Warriner et al., 2013). However, only very few researchers have investigated semantic prosody with the help of such datasets (Dilts & Newman, 2006; Snefjella & Kuperman, 2016; Winter, 2016). Doing so advances the reproducibility of semantic prosody research because given the same corpus and the same valence dataset, different researchers will come to the same conclusions.

Moreover, a reproducible workflow would allow semantic prosody research to address a common critique raised against it—namely, that the field has overly focused on a handful of isolated examples, such as *set in* and *cause* (see Stewart, 2010; Whitsitt, 2005). The manual interpretation of concordance lines impedes research progress in the domain of semantic prosody because it means that the analysis of many linguistic examples cannot be streamlined, precluding general statements about the vocabulary of English.

10.5. A manifesto for norms

Given problematic cases such as the above-mentioned examples of synesthetic metaphors and semantic prosody, what are we to do? Paradoxically perhaps, my solution to the abundance of introspection in linguistics will be to incorporate even more introspection into linguistic analysis. However, I will do so differently, using rating studies.

To justify my approach, it is important to start with the existing criticism of introspective judgments. Much ink has been spilled about the role of intuition in the language sciences, and researchers have rightly pointed out the limitations of introspective judgments, especially when the conclusions of entire theories rest on the intuitions of individual linguists.[2]

The problem of overly relying on individual intuitions has been explored extensively in the literature on grammaticality and acceptability judgments. To give just a few examples, experiments have shown that grammaticality judgments by linguists deviate from those of nonlinguists (Spencer, 1973; see also Dąbrowska, 2010). Linguistic judgments change further through repeated exposure to particular sentences, as well as through sentence context (Nagata, 1988, 1989). There are also anchoring effects (Nagata, 1992), where the perceived grammaticality of a sentence depends on whether it was presented alongside ungrammatical sentences. Perhaps one of the most pressing problems is theoretical biases: Schütze (1996, p. 187) rightly says that "one cannot find a more biased subject than the investigator" (see also Dąbrowska, 2016b, p. 57; Gibbs, 2007, pp. 3–4; see also Gibson & Federenko, 2010).

Many researchers have discussed or problematized the role of intuitive judgments in the language sciences, for many more reasons than the ones just listed (Carrol, Bever, & Pollack, 1981; Cowart, 1997; Dąbrowska, 2010, 2016a, 2016b; Edelman & Christiansen, 2003; Featherston, 2007; Ferreira, 2005; Gibbs & Clark, 2012; Gibson & Fedorenko, 2013; Marantz, 2005; Myers, 2009; Pullum, 2007; Schütze, 1996; Sprouse & Hornstein, 2013; Wasow & Arnold, 2005). Yet, many researchers also believe that it is impossible to completely disband the role of intuitions when studying language. Gries and Divjak (2010, p. 336) state that "no scientist in the humanities or social scientists would deny that some degree of intuition plays a role in nearly *any* study." This is particularly the case when the *content* of conceptualization needs to be probed. As stated by Dąbrowska (2016b, p. 56), introspective judgments "provide the most direct source of information about some aspects of language, notably meaning."

This is where rating studies enter. The term "norms" is frequently used by psycholinguists to refer to rating data that is subsequently used in experiments (it should be kept in mind that there is nothing "normative" about norms). Rather than relying on a single linguist's intuition about a linguistic construct, we are better off aggregating over many intuitions from laypeople. By using norms, intuitions are treated more carefully, using the wisdom of the crowd. This also addresses Dąbrowska's (2016a) plea for cognitive linguists to pay more attention to individual differences. There is an increasing amount of evidence for people differing in their linguistic and conceptual systems (Dąbrowska, 2012, 2015). Precisely because of this, it is necessary to aggregate over the responses from different individuals when trying to achieve highly generalizable claims. We cannot just rely on just one individual's intuitions, especially if that individual is a linguist.

Norms generated by native speaker ratings have been around for a long time (e.g., Osgood, Suci, & Tannenbaum, 1957; Paivio, Yuille, & Madigan, 1968), but recently there has been an influx of massive studies that normed thousands of English words for many different dimensions, including iconicity (Perry, Perlman, & Lupyan, 2015; Perry, Perlman, Winter, Massaro, & Lupyan, 2018), age of acquisition (Kuperman et al., 2012), concreteness (Brysbaert, Warriner, & Kuperman, 2014), emotional valence (Warriner et al., 2013), sensory modality (Connell, Lynott, & Banks, 2018; Lynott & Connell, 2009, 2013; Speed & Majid, 2017; van Dantzig et al., 2011; Winter, 2016), as well as more specific dimensions such as a word's association with pain, manipulability, color, sound, and motion (Amsel, Urbach, & Kutas, 2012; Medler et al., 2005), or whether a word describes rough and smooth surfaces (Stadtlander & Murdoch, 2000). Many of these norms were collected with psycholinguistic experiments in mind. However, nothing prevents using them together with corpora (Chapters 14, 15, & 17), or studying norms in their own right (e.g., Chapters 12–13).

I believe that the potential for norms is not recognized enough by linguists, particularly by cognitive linguists and corpus linguists. Here is a list of key advantages of a norm-based approach:

- Norms are collected from hundreds of individuals; because people differ in their judgments and their understanding of language (Dąbrowska, 2012, 2015), aggregation over multiple ratings affords better generalization of common patterns.

- Norms are collected from naïve language users and thus help to reduce bias from linguistic theories (e.g., Dąbrowska, 2016a, 2016b, p. 57; Gibbs, 2007; Schütze, 1996, p. 187; Spencer, 1973).

- Norms are collected before an actual investigation takes place, independent of that investigation (e.g., a particular corpus analysis); as such, the collection of norms cannot be biased by the research question at hand; this also means that there can be no ad-hoc additions or subtractions of linguistic items to or from datasets, such as may occur when manual corpus annotation is performed.

- To allow the collection of norms, participants need precise instructions (e.g., what is it exactly that should be rated?); this forces researchers to operationally define their linguistic constructs more explicitly.

- Depending on how the rating task was set up, norms make continuous data available—for example, words can be rated on a scale from positive to negative (e.g., Warriner et al., 2013); because meaning is flexible and graded (e.g., Croft & Cruse, 2004), this captures shades of meaning more accurately.

- The numerical format of linguistic norms makes additional statistical analyses possible (see Chapters 12–17)

- A norm set collected for one study can be used later for another analysis; sharing norm sets within the community streamlines future research and also allows comparisons across studies; through this, norms allow to build "sufficient common ground for a stringent comparison of competing models" (Grondelaers, Geeraerts, & Speelman, 2007, p. 167)

- Norms provide a means out of the circularity trap that haunts much of cognitive linguistics; that is, one cannot argue for conceptual explanations of linguistic patterns based on linguistic evidence alone (Dąbrowska, 2016a; Gibbs, 2007; Murphy, 1996, 1997); since norms are collected separate from an analysis of linguistic patterns (such as of corpus data), they provide an alternative access to conceptualization.

Thus, there are many advantages to using norms. At a bare minimum, it must be acknowledged that norms are better than what they are supposed to replace: the subjective judgments of individual linguists with theoretical predispositions.

However, it has to be kept in mind that norms are still introspective judgments performed on linguistic items. Norms are, after all, subjective; they are a form of self-report. Time and time again, psychological research has shown that self-report can differ substantially from actual behavior (e.g., Nisbett & Wilson, 1997). Spence and Piqueras-Fiszman (2014) rightly remark that "as psychologists, we are highly reticent about putting too much weight on the unconstrained self-report of participants" (pp. 18–19). Because of this, it has to be kept in mind that the judgments we collect are metalinguistic, differing from actual language use. This was actually already discussed in Chapter 7, where I argued that when the word *sweet* is seen in isolation, it is judged to be primarily a taste word, even though it could be a smell word in actual use. The judgments that are included in norms are decontextualized, whereas actual language use always happens in context. Some of these limitations can be dealt with by re-contextualizing norms in combination with corpora, as will be done in Chapters 14, 15, and 17.

To conclude, norms are essential for the language sciences, and compared to the high preponderance of work relying on introspective judgments of individual linguists, norms are under-utilized. Hunston (2002, p. 22) states that "although an over-reliance on intuition can be criticised, it would be incorrect to argue that intuition is not important." This chapter argues for making intuitions take a position within linguistic research that is even more center stage, but to do so in a way that is more reliable and more scalable.

10.6. "Fuck nuance"?

How do norms fit with the theoretical demands of the language sciences? In his programmatic article "*Fuck nuance*," Healy (2017) outlines what he calls "nuance traps" in theory-building. Healy criticizes that many social scientists tend to make their theories ever more complicated by incorporating ever more nuanced aspects of a studied phenomenon:

> By calling for a theory to be more comprehensive, or for an explanation to include additional dimensions, or for a concept to become more flexible

and multifaceted, we paradoxically end up with less clarity. We lose information by adding detail. (2017, p. 122)

Healy (2017, p. 123) also calls for abstraction, saying that "Theory is founded on abstraction, abstraction means throwing away detail for the sake of a bit of generality, and so things in the world are always 'more complicated than that'—for any value of 'that'." Ever more complex theories become ever more difficult to understand, and ever more complex theories become ever more difficult to falsify. Complexity is certainly needed in some cases, but it can also work against generalizability in others. We want generalizable theories, and we want theories that are easy to understand. We sometimes need to shed the complexities of a study topic to reach higher levels of abstraction.

In semantic research, the incorporation of norms achieves such abstraction. Each rating study operationalizes only one aspect of meaning, deliberately ignoring everything else. As characterized by Dienes (2008, p. 3), "an operational definition defines the meaning of a concept in terms of the precise procedures used to determine its presence and quantity." Thus, the analyst may carve up the lexicon into dimensions of interest, each dimension being precisely defined, with ratings for each dimension ascertained in a rigorous and reproducible manner. By collecting more norms for more dimensions, linguistics can begin to get a deeper understanding of language.

The above-mentioned research on semantic prosody is a great example of how more nuance may impede progress in theorizing. It is clear that the notion of semantic prosody cannot be fully reduced to the positive/negative dimension (Bednarek, 2008; Hunston, 2007). However, part of the problem of semantic prosody research is that it has perhaps paid too much attention to the nuanced shades of meaning of specific examples, at the expense of characterizing large chunks of the lexicon in more abstract and generalizable terms.

In Chapter 16, I will operationally define semantic prosody in terms of the positive/negative dimension (compare Morley & Partington, 2009), using the emotional valence norms from Warriner et al. (2013). The pay-off of this approach is that the analysis of semantic prosody becomes reproducible (and less open to such attacks as those by Whitsitt, 2005, and Stewart, 2010). Moreover, using a quantitative approach that deliberately simplifies the complex world of human emotions means hundreds of words can be analyzed in an automated fashion (see also Snefjella & Kuperman, 2016), thus yielding highly generalizable results. A final advantage of this approach is that the particular view of semantic prosody endorsed here is more precisely specified thanks to being operationalized via the valence scale.

The points Healy (2017) makes also apply to sensory words. As discussed in Chapter 2 and Chapter 7, I am highly critical of blind applications of the five senses folk to perceptual language. Nonetheless, I am adopting the five senses folk model in the analyses that follow. This is because lumping the senses together into five big categories (rather than splitting the sensory world into many smaller subcategories) allows for making highly general claims with a high degree of abstraction. Moreover, in Chapter 11, I will present a dataset in which the five senses are precisely specified in terms of five numerical scales. This precise operationalization actually makes it possible to criticize the five senses folk model with empirical data. Thus, temporarily adopting the five senses folk model allows falsification. Sometimes it is best to operate within a theory to prove it wrong.

10.7. Comparison to other approaches in empirical semantics

Any methodological choice needs to be defended against other choices that could have been made. Therefore, in this section I will look at other ways through which sensory meaning could be quantified. Specifically, what other approaches allow studying sensory meaning empirically? Here, I want to compare the norm-based approach to two alternative approaches to measuring meaning: first, experimental approaches within the literature on embodied cognition, and second, distributional semantics.

As reviewed in Chapter 5, there is by now a large body of evidence showing that understanding language involves some form of perceptual, motoric, or emotional simulation. According to simulation-based accounts, performing a mental simulation of the content of an utterance *is* meaning (Bergen, 2012). Take, for example, a study by Zwaan, Stanfield, and Yaxley (2002) who looked at how language users mentally simulate visual shapes. These researchers showed that processing an image of an eagle with its wings extended (rather than tucked in) was faster after reading the sentence *The ranger saw the eagle in the sky* compared to the sentence *The range saw the eagle in the nest*. By pairing nonlinguistic visual stimuli with linguistic stimuli, this experiment avoids the circularity trap of trying to infer conceptual content directly from linguistic patterns (see discussion in Dąbrowska, 2016a; Gibbs, 2007).

However, experimental approaches to word meaning are limited in their scalability when it comes to making generalizations about large-scale linguistic patterns, such as the structure of the mental lexicon. The just-mentioned study on the perceptual simulation of shape characteristics assessed a total of 24 sentence-picture pairs (Zwaan et al., 2002), a figure that is representative of this literature. The experiment is designed to assess a particular theory of linguistic processing; the approach cannot easily be extended to large numbers of sentences or words. Running experiments on so many linguistic stimuli is simply infeasible. On the other hand, the norm-based approach has impressive scalability. Norms are generated from quick judgments on linguistic stimuli and can hence be collected for thousands of words, as attested by recent so-called "megastudies" such as the concreteness norms for ~40,000 English words by Brysbaert et al. (2014) and the emotional valence norms for ~14,000 English words by Warriner et al. (2013).

Scalability is also afforded by distributional-semantic approaches to meaning. These approaches infer meaning from language use, such as whether words occur together in the same contexts. For instance, one could infer that *nurse* is semantically related to *doctor* by looking at whether these two words co-occur together, or whether they occur in similar types of linguistic contexts (e.g., Miller & Charles, 1991). A large amount of work in corpus linguistics, computational linguistics, and natural language processing attests to the usefulness of inferring meaning from contexts. However, as stated by Dąbrowska (2016a, p. 489), "while distributional features provide important clues to meaning, adequate semantic descriptions must make use of other methods as well." Distributional approaches can easily yield reliable estimates of the semantic similarity of such words as *doctor* and *nurse*, but they do not tell us what these words ultimately *mean*.

In contrast to distributional approaches, norms provide direct access to what words mean, albeit only to highly circumscribed parts of meaning, such as whether the meaning is concrete or not (Brysbaert et al., 2014), whether it is positive or not (Warriner et al., 2013), or how the word relates to the five senses (Lynott & Connell, 2009). Without any external input, distributional semantics is limited to quantifying semantic similarity between linguistic items, rather than any features that directly relate to the item's meaning itself.

Neither psycholinguistic experiments, nor distributional semantics, nor a norm-based approach can be considered optimal with respect to the study of meaning. Instead, each approach has its own set of advantages and disadvantages. Moreover, each approach is used to answer different kinds of questions. When it comes to studying perceptual language, however, norms provide an ideal way of forming a big-picture view of the sensory lexicon of a language, such as English. It should also be pointed out that the three approaches outlined here can be combined in fruitful ways. This will be done in Chapters 14 and 17, where I use norms to look at word usage in context, investigating patterns of co-occurrence between different types of sensory words.

Besides making novel theoretical contributions, this book shows how far we can take the norm-based approach.

10.8. Limitations

Any investigation should be clear about its limitations. The analyses that follow this chapter focus on sensory adjectives at the expense of considering other parts of speech, such as nouns, verbs, or ideophones. Moreover, the following analyses are focused on synchronic data only. Although there is clearly a historical dimension to sensory language (e.g., Williams, 1976), the analysis of diachronic data will have to be left for future studies. Nevertheless, it must be emphasized that the methods introduced in the following chapters are very useful for quantitative historic analyses as well.

My main focus will be on investigating the highly conventionalized part of the sensory vocabulary in a fairly abstract manner, neglecting how sensory words are used in particular discourse contexts. At present, my aim is not to provide a detailed qualitative description of how sensory words are used in situated interactions or specific types of discourse;I want to characterize the broad characteristics of the sensory vocabulary. It is hoped that this “big picture” of the English sensory vocabulary facilitates future qualitative research in the domain of sensory language.

Throughout all empirical chapters of this book (Chapters 11-17), I will adopt a "cultural tool" perspective, focusing on the *inventory* of sensory words and how this inventory is deployed in use. Clark (1996, p. 56) distinguishes between the product tradition of studying language and the action tradition of studying language. Whereas the former focuses on language as a set of entities, the latter focuses on real cases of language use. An important difference between the product and the action tradition is that the product tradition is confined to an analysis of *potential* uses, whereas the action tradition looks at *actual* uses. In this book, we will both look at potential and actual uses, but our analysis of actual uses is only aimed at making statements *about* the toolkit of sensory words. It clearly is important to consider language as a form of action (Clark, 1996; Wilce, 2009: 20-21), but this will not be the focus adopted here.

Another caveat has to do with the focus on individual words as the units of analysis. This focus should not be taken to suggest that I endorse a fully lexically based model of language. Many corpus linguists correctly point out that the real units of speech and writing are not individual words, but larger-than-word units (Sinclair, 1996; Stubbs, 2001), or "patterns" (Hunston & Francis, 2000). However, to understand how sensory language is employed in context, isolated sensory words such as *spotted*, *amber*, *glittery*, *tangy*, *smooth*, and *prickly* form an ideal starting point to investigate what linguistic resources for expressing perceptual content the English language makes available to its speakers.

The final limitation is that the analyses presented in Part II exclusively deal with English. Future work needs to conduct similar analyses in other languages. The methods introduced in the following chapters can easily be extended to the study of other languages.

In short, I will only focus on English, only focus on adjectives, and only on the present. However, despite these limitations, the analyses presented in the following chapters deal with foundational issues in sensory linguistics, such as how the senses are connected in the lexicon and how sensory language is used for evaluative purposes. Many of the topics I will cover are also relevant to crosslinguistic or historical analyses, and to the analyses of other types of word classes, such as perception verbs. To this extent, the findings reported in the following chapters are relevant to *all* of sensory linguistics.

10.9. Statistics

Because each chapter uses very different statistical techniques, explanations for particular approaches will be given within each chapter. For all chapters with data (Chapters 12–17), the following GitHub repository contains analysis code and raw data:

https://github.com/bodowinter/sensory_linguistics

The R statistical programming environment version 3.3.1. was used for all analyses (R Core Team, 2016). The `tidyverse` packages (Version 1.1.1.) were used for data processing (Wickham, 2017b) and the `stringr` package version 1.2.0 was used for character processing (Wickham, 2017a). The packages `GISTools` 0.7-4. (Brunsdon & Chen, 2014), `png` 0.1-7. (Urbanek, 2013), and `plotrix` 3.6.5. (Lemon, 2006) were used for visualization. Hartigan's dip test (Chapter 12) was performed with the `diptest` package version 0.75-7 (Maechler, 2015). Gaussian mixture models (Chapter 13) were performed with the `mclust` package version 5.2 (Fraley, Raftery, Murphy, & Scrucca, 2014). Linear mixed effects models (Chapter 17) were computed with the `lme4` package version 1.15.6 (Bates, Maechler, Bolker, & Walker, 2015a, 2015b). Negative binomial regressions (Chapter 14) were computed with the `MASS` package version 7.3.45 (Venables & Ripley, 2002). Variance inflation factors (Chapters 15–16) were computed with the `car` package version 2.1.3 (Fox & Weisberg, 2011).

The sense icons used throughout the figures of this book are taken from freepik.com: https://www.freepik.com/free-vector/five-senses-icons_837465.htm

---

[1] Ignoring the type/token distinction characterizes work on synesthetic metaphors to this day (but see Ronga, 2016). For example, Shen (1997) lists counts that are based on tokens (Shen, personal communication), but this is not stated as such in his paper.

[2] Stewart (2010) rightly points out that intuition and introspection are not the same thing. However, for present purposes, this distinction is irrelevant.

Chapter 11. Norming the senses

11.1. Classifying sensory words

Other works have focused on how speakers of different languages talk about the act of perception, as is the case when speakers use verbs such as *to see*, *to hear*, *to feel*, *to taste*, and *to smell* (Evans & Wilkins, 2000; Nakagawa, 2012; Sweetser, 1990; Viberg, 1983). Languages have dedicated communicative tools to talk about perception as an activity, something that humans *do*. However, languages also have dedicated means to talk about the results of this activity, the content of perception. Part II of this book focuses on how the English language expresses perceptual content, rather than perceptual activity. This motivates the focus on sensory adjectives, which are specifically *about* sensory properties, such as color, texture, and taste.

To study sensory language empirically, one first needs to assemble a sizeable list of sensory terms (Ronga, 2016; Ronga et al., 2012; Strik Lievers, 2015). Lehrer (1978) already mentioned that for investigating language and perception, "what is needed is a study of something *like* a random sample of lexical items" (p. 95, emphasis in original). When doing sensory linguistics, a random sample of words may not actually be what we want. A random sample of words might include many words that are not strongly related to the senses at all, such as *freedom* and *governance*. Instead, we want a word list that allows us to address questions specifically relating to the senses, which means that we have to zoom in on perceptual vocabulary. More important than obtaining a random sample of words, a word list must be assembled independently from the theory one wants to test (see Chapter 10). Only by having an unbiased word list of sensory words can one confirm or disconfirm a theory about language and perception.

To study differences between the five senses, the word list furthermore needs to be classified according to some perceptual dimension, such as to which sensory modality each word relates. This latter step is complicated by the already mentioned fact that some sensory words are highly multisensory (Diederich, 2015; Fenko et al., 2010; Lynott & Connell, 2009; Paradis & Eeg-Olofsson, 2013); that is, they evoke more than just one sensory modality. For example, how are we to classify adjectives such as *barbecued* and *fishy*? Do these terms describe perceptual properties? And if so, are they gustatory, olfactory, or both? What about the word *crunchy* (cf. Diederich, 2015), which is treated as auditory in Ronga (2016, p. 57)? Does this word not equally well describe something that is also tactile, and perhaps also gustatory? Finally, what about common sensibles (Chapter 4.3.2)? (see Levinson & Majid, 2014, pp. 412–413; Marks, 1978, Chapter 2; Ronga et al., 2012, pp. 149–150; Sorabji, 1971). Tekiroğlu, Özbal and Strapparava (2014, p. 1) state that "connecting words with senses…is a straightforward task for humans by using commonsense knowledge." The existence of multisensory words (such as *harsh*) and words describing common sensibles (such as *large*) shows that this is not so straightforward a task at all.

This book advocates for the use of modality norms in sensory linguistics. There are by now many rating studies that have collected such norms, such as Lynott and Connell (2009, 2013), van Dantzig et al. (2011), Winter (2016), and Speed and Majid (2017).

The first researchers to collect sensory modality norms for all five senses were Lynott and Connell (2009).[1] These researchers asked 55 native speakers of British English to rate how much a given property is experienced "by seeing," "by hearing," "by feeling through touch," "by smelling" and "by tasting." For each of the modalities, participants responded on a scale from 0 to 5. This not only allows quantifying the degree to which a word corresponds to a specific sense, but, as we will see shortly, it also offers ways of quantifying the degree to which a word is multisensory. Notice a crucial assumption of the Lynott and

Connell (2009) rating study: Modality association is a continuous quality. That is, a word can be more or less visual, more or less auditory, and so on.

Lynott and Connell (2009) collected norms for a total of 423 object properties. The list was assembled partly based on previous research on sensory words and partly based on dictionary searches. The adjectives were selected with psycholinguistic experiments in mind. Importantly, the list is not a random sample of sensory words.

Table 4 shows two example words, *yellow* and *harsh*, with their corresponding average perceptual strength ratings (one mean averaged over all participants per modality).[2] The rightmost column specifies each word's so-called "modality exclusivity." This measure is defined as the range of perceptual strength ratings divided by the sum. The resulting proportion is then multiplied by 100 so that exclusivities can be expressed as percentages. An exclusivity of 0% would mean that a word has exactly the same rating for all sensory modalities—that is, it is maximally multisensory. An exclusivity of 100% represents the maximum possible unisensoriality—only one sense is rated above zero. The word *yellow* has a considerably higher exclusivity (95.1%) because the only modality rated as very high in perceptual strength was the visual one. The word *harsh* has a much lower exclusivity (11.6%) because the perceptual strength ratings are distributed across all five senses. The fact that continuous perceptual strength ratings enable the calculation of the modality exclusivity measure shows the utility of having a numerical scale; it allows dealing with the multisensoriality of perceptual vocabulary in a principled manner.

Table 4

Modality norms for 'yellow' and 'harsh'

	Visual	Tactile	Auditory	Gustatory	Olfactory	Exclusivity
yellow	**4.9**	0.0	0.2	0.1	0.1	95.1%
harsh	3.2	2.5	**3.3**	2.3	1.8	11.6%

Figure 2 shows the distribution of modality exclusivity scores. The most unisensory adjective in the Lynott and Connell (2009) dataset is *brunette* (98%); the most multisensory word is *strange* (10%). The average modality exclusivity is 46%, which indicates that many adjectives are multisensory. This fact alone should be a reason to consult modality norms rather than relying on hard-cut sensory classifications, where a word can only be associated with one sensory modality (see discussion in Lynott and Connell, 2009).

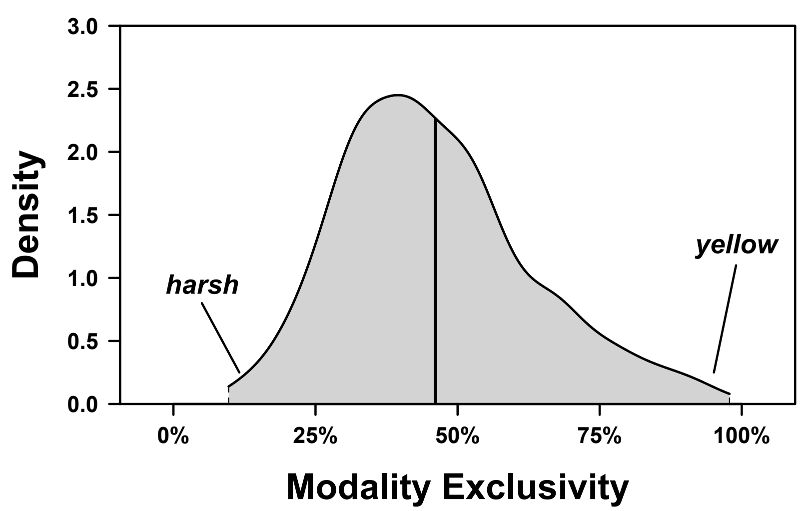


Figure 2. Kernel density estimates of modality exclusivity ratings for adjective norms. Data shows 423 adjectives from Lynott and Connell (2009); the x-axis represents the modality exclusivity with the least unisensory words to the left and the most multisensory words to the right; the y-axis represents how many

words there are for a given exclusivity; density curves are restricted to the observed range; solid vertical line indicates the mean exclusivity of 46%.

The highest perceptual strength ratings in their respective categories were assigned to the words *bright* (sight), *barking* (sound), *smooth* (touch), *citrusy* (taste), and *fragrant* (sound). A word's highest perceptual strength rating determines what Lynott and Connell (2009) call "dominant modality." In Table 4, *yellow* is classified as visual because its visual strength rating is higher than its other perceptual strength ratings. The word *harsh* is classified as auditory because its maximum perceptual strength rating belongs to the auditory modality. However, the word barely passes as an auditory word: Vision has almost the same perceptual strength rating, and the other senses receive relatively high ratings for the word *harsh* as well. The contrast between the adjectives *yellow* and *harsh* clearly shows that classifying words according to dominant modalities is inherently more meaningful for words that are relatively more unisensory. Because of their difference in modality exclusivity, the classification of *yellow* as visual appears to be more adequate than the classification of *harsh* as auditory.

For each dominant modality, Table 5 lists the most and the least frequent words, as well as the most and the least exclusive words (frequency data was taken from the SUBTLEX movie subtitles corpus, Brysbaert & New, 2009).

Table 5

Example adjectives by sensory modality

Modality	**Frequent**	**Infrequent**	**Unisensory**	**Multisensory**
Visual	*little, big*	*vegetal, craggy*	*brunette*	*strange*

Tactile	*hard, cool*	*bristly, brackish**	*stinging*	*brackish**
Auditory	*quiet, crying*	*soundless, sonorous*	*echoing*	*harsh*
Gustatory	*sweet, delicious*	*honeyed, orangey*	*bitter*	*mild*
Olfactory	*fresh, burning*	*burnt, reeking*	*perfumed*	*burning*

Note. The two most frequent and infrequent adjectives for each sensory modality based on the SUBTLEX movie subtitle corpus (Brysbaert & New, 2009) and the most and least exclusive adjective; data from Lynott and Connell (2009); *the word *brackish* has been classified as predominantly tactile by the participants of this study even though dictionaries commonly define this word as "slightly salty".

In a second norming study, Lynott and Connell (2013) collected perceptual strength ratings from 34 native speakers of British English for a set of 400 randomly sampled nouns. Although the focus here is on sensory adjectives, I must also discuss noun norms. These are important as a point of comparison to the adjective norms, and they will become important when looking at adjective–noun pairs in Chapters 14 and 17. Table 6 gives several examples from the norming study of Lynott and Connell (2013).

Table 6

Example nouns by sensory modality

Modality	**Frequent**	**Infrequent**	**Unisensory**	**Multisensory**
Visual	*want (n.), back (n.)*	*provision, builder*	*reflection*	*quality*
Tactile	*feel (n.), hold (n.)*	*strain, item*	*hold (n.)*	*item*
Auditory	*help (n.), God*	*socialist, quarrel*	*sound*	*heaven*
Gustatory	*honey, food*	*sauce, supper*	*taste (n.)*	*treat (n.)*

Olfactory	*air, breath*	-	-	-

Note. The two most frequent and infrequent nouns for each sensory modality (based on SUBTLEX) and the most and least exclusive noun; data from Lynott and Connell (2013).

Among the nouns, smell is severely underrepresented, with only two nouns (*air* and *breath*). Moreover, these nouns do not directly relate to smell so as much as they relate to the sensory organs of the nose and the mouth. With an average exclusivity of 39%, the nouns are more multisensory than the adjectives (46%), a difference that is statistically reliable (unpaired Wilcoxon rank sum test: W = 103270, p < 0.0001). Lynott and Connell (2013) argue that this is because nouns are used to refer to objects, events, and actions, which can generally be perceived through multiple modalities. For example, food can readily be seen, smelled, and tasted. Adjectives, on the other hand, highlight specific properties of objects and actions, and as such, they are more likely to single out specific content from a particular modality (e.g., Givón, 2001, p. 53). Whereas the noun *food* is highly multisensory (18% exclusivity), the expressions *fragrant food* and *tasty food* highlight modality-specific sensory aspects of the food. The observation that nouns are more multisensory than adjectives fits Gärdenfors' (2014) "general single-domain thesis" according to which "words in all content word classes, *except for nouns*, refer to a single domain" (p. 239, emphasis added).

Another potential reason for the lower exclusivity scores of the noun norms from Lynott and Connell (2013) might have to do with abstractness: Nouns such as *information*, *fact*, and *socialist* denote concepts that cannot easily be experienced directly through any of the senses. With these highly abstract concepts, the dominant modality classification is often questionable; indeed, it is overall questionable whether perceptual strength ratings are appropriate for such concepts. For instance, the noun *welfare* is listed in Lynott and Connell (2013) as having vision as its dominant modality, but this word received relatively low perceptual strength ratings overall. Because it is not a very sensory word in the first place, the question as to which modality it belongs does not carry much weight.

One has to be careful, however, in comparing the noun and adjective norms. The nouns were randomly sampled (Lynott & Connell, 2013), but the adjectives were not. This difference in how the word lists were created may affect the results: Winter (2016) collected perceptual strength ratings for verbs using two lists: one compiled using a dictionary search procedure based on typical perception verbs (*to see, to hear* etc.), and the other compiled using random sampling. The manually constructed list had higher exclusivity (57%) than the random sample (44%). Thus, the exclusivity difference between the Lynott and Connell (2009) adjectives and the Lynott and Connell (2013) nouns cannot be taken at face value. In particular, precisely because the adjective list in Lynott and Connell (2009) was constructed with psycholinguistic experiments on sensory language in mind, the words are expected to be high in specific sensory content relative to many other adjectives that could have been in the dataset, such as *stupid*, *intelligent*, *rich*, and *poor*. Thus, it is not entirely clear at present whether the lower modality exclusivity of the nouns is indeed due to a difference in lexical category, or due to a difference in sampling.

Although I strongly advocate for using modality norms in sensory linguistics, it should be noted that the present datasets are not without their flaws. Some problems include straightforward misunderstandings of the meaning of particular words. For example, *firm* (n.) in the Lynott and Connell (2013) noun rating study received the highest perceptual strength rating for the tactile modality, presumably because participants were not thinking of the noun *firm* (as in meaning 'company') but of the adjective *firm*, which relates more directly to a felt impression. In Lynott and Connell (2009), participants rated *clamorous* to be higher in tactile strength (2.9) than in auditory strength (2.4), even though most dictionary definitions emphasize the auditory meaning of this word. Another problematic classification is *brackish* (see Table 5 above), which had a higher tactile strength (3.0) than gustatory strength (1.5) and olfactory strength (1.5), despite dictionary definitions being about salty water, which is arguably more related to taste. It is noteworthy that both *clamorous* and *brackish* are very infrequent words (SUBTLEX frequencies 0 and 4 respectively), which suggests that the meaning of these words was perhaps not sufficiently known to the participants of the Lynott and Connell (2009) rating study. However, all in all, these minor problems pose no threat to the conclusions reported in this book: A few isolated cases are unlikely to skew a result that is based on a total set of 423 words.

11.2. Avoiding circularity

A bigger methodological issue has to do with the following question: How did participants actually perform the modality rating task—that is, on what were the participants in Lynott and Connell (2009) basing their judgments? Remember that participants were asked how much a given property, say *yellow*, was experienced "through vision" or "through hearing" and so on. In simple cases of making judgments on unisensory words this appears to be straightforward: *Yellow* is clearly visual. But in the case of relatively more multisensory words, how did participants decide how each modality should be rated? One likely strategy that participants may have adopted is to generate linguistic examples:

For instance, to determine what the modality of *harsh* should be, a participant may have thought of linguistic examples such as *harsh sound* or *harsh taste*.

If such a linguistic strategy had been adopted by the participants in Lynott and Connell (2009), the modality norms would be problematic when combined with corpora. For instance, Chapter 14 finds that touch ratings and sight ratings correlate with each other in corpus data. However, if the ratings are themselves influenced by language, then this finding would be circular, perhaps even inevitable. Simply put, if participants used a context-retrieval strategy when performing the rating task in Lynott and Connell (2009), then these ratings cannot be used to also analyze linguistic contexts in corpora.

A modality norming study conducted by van Dantzig and colleagues (2011) can be used to address this concern. In this study, each property label was presented in the context of two objects. For instance, the property described by the word *abrasive* was assessed as it pertains to *lava* and *sandpaper*. Participants then provided perceptual strength ratings, as in the Lynott and Connell (2009) study. Pairing adjectives with nouns gives participants specific examples to consider, thus binding their property ratings to particular objects. This presumably restricts the influence of spontaneously generated linguistic examples.

Finding that the results from van Dantzig et al. (2011), which had contexts, are similar to the results from Lynott and Connell (2009), which had no contexts, would alleviate the concern of circularity. For the 365 words that are overlapping between the two datasets, 82% share the same dominant modality classification. Moreover, the mean perceptual strength ratings of the two studies correlate reliably (all p's < 0.05) with very high Pearson's correlation coefficients, ranging from $r = 0.81$ for vision to $r = 0.92$ for audition.[3] The modality exclusivity scores from the Lynott and Connell (2009) dataset also correlate reliably and very strongly with the van Dantzig et al. (2011) exclusivities, with r = 0.75 (t(363) = 21.69, $p < 0.0001$).

To compare the degree to which words present in both lists are similar overall (across all five modalities), we may compute the cosine similarity between the modality vectors of each word. Each word in the modality norm dataset can be represented as a vector of five numerical values, one for each sensory modality. The word *abrasive* for example, can be represented as A = [2.9, 3.7, 1.7, 0.6, 0.6] (sight, touch, sound, taste, smell). This vector encapsulates what one may call the "modality profile" of this word. The word *abrasive* also has a modality vector in the van Dantzig et al. (2011) dataset, which turns out to be B = [3.6, 3.6, 1.8, 0.3, 0.2]. The two vectors for the word *abrasive* are highly similar to each other: Modalities that have relatively high numbers in one rating study also tend to have high numbers in the other; and vice versa for low numbers. For two vectors A and B, the cosine similarity is defined as follows:

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \qquad \text{(E1)}$$

The resulting metric ranges from 0 to 1. If the adjectives from two datasets are exactly the same with respect to all five senses, the cosine will be exactly 1 (maximally similar). Maximally different adjectives would have a value of 0. The average cosine similarity for the words present in both datasets is 0.97. The cosine values ranged from 0.74 to 0.99. This indicates a very high fit between the Lynott and Connell (2009) ratings and the van Dantzig et al. (2011) ratings. Throughout this book, the Lynott and Connell (2009) dataset will be used because it is more extensive than the van Dantzig et al. (2011) dataset (423 as opposed to 387 words).

11.3. Comparison to other approaches

Using norms generated by native speakers is not the only method for generating a word list of sensory terms. To explain my reason for using modality norms as opposed to other methods, a comparison to other approaches is necessary.

Many researchers use dictionaries to generate a list of sensory terms (e.g., Bhushan, Rao, & Lohse, 1999; Ronga et al., 2012; Strik Lievers, 2015). With this approach, a set of seed words that appear to clearly correspond to a particular modality is selected, such as the verb *to hear* for audition. Then, this initial set is enlarged by considering all the synonyms of the seed words. For example, the Collins Dictionary lists *to listen to* and *to eavesdrop* as synonyms of *to hear*. Thus, *eavesdrop* and *listen* are added to the list of auditory terms. For this approach to always yield a trustworthy modality classification, synonyms of a perceptual word from one particular sensory modality need to always be from the same sensory modality. However, this clearly is often not the case. For instance, Collins lists *to attend to* as a synonym of *to hear*, even though this word does not actually strongly relate to the auditory modality—one can attend to the subjective impression of taste and smell just as much as one can attend to a sound. An even more problematic example is the fact that *to perceive* comes up as a synonym of most perception verbs. In general, the thesaurus-based approach always needs an additional step of modality classification because not all words unequivocally belong to a particular modality.

Strik Lievers (2015) used dictionaries and thesaurus lists to compile a dataset of sensory words, which she subsequently hand-annotated for the particular sensory modality involved. We can compare her list to the native speaker ratings from Lynott and Connell (2009). For those 96 words represented in both datasets, 92% of them have the same modality classification in both.

Some differences in classification include *burning* and *pungent*, which are classified as haptic/tactile in Strik Lievers (2015), but as being predominantly olfactory in the modality norms. Similarly, *acrid*, *aromatic*, and *smoky* are classified as gustatory in Strik Lievers (2015), but as predominantly olfactory in the study by Lynott and Connell (2009). I am not stating that either one of these classifications is more correct than the other. However, in contrast to Strik Lievers' word list, the modality norms allow dealing with multisensoriality in a more principled fashion. In the Lynott and Connell (2009) norms, words such as *aromatic* and *pungent* are treated as belonging to multiple sensory modalities. Moreover, the modality exclusivity measure itself makes interesting data available that can be studied in its own right to investigate the properties of the sensory lexicon (Chapter 12). Because of the lack of a continuous rating scale, the multisensoriality of sensory words, as quantified through exclusivity, cannot be studied with the Strik Lievers (2015) word list.

Tekiroğlu and colleagues (2014) constructed a word list in an automatic fashion using various techniques from natural language processing that were applied to a range of sources, including FrameNet, WordNet, and the GigaWord corpus. Because the approach was automatic and did not rely on tedious (and expensive) human judgment, their list is much larger than any other sensory word list so far, containing 22,684 entries. The core idea behind the Sensicon's way of classifying the senses is that a word's meaning can be inferred through the types of contexts in which a word occurs. For example, we may expect that taste words occur together with other taste words, such as when a speaker or writer talks about food. The Sensicon considers a local context of four content words before and after the candidate word.

The resulting dataset, however, has several odd classifications. For instance, the three most visual adjectives are indicated to be *federal*, *large*, and *blue*; the three least visual adjectives are *salty*, *teen*, and *sour*. What exactly is visual about the word *federal*, in particular when compared to the allegedly non-visual word *teen*? The three most olfactory adjectives in this dataset are *rich*, *musky*, and *hot*; the three least olfactory "adjectives" (using the part-of-speech labels provided within the Sensicon) are *federal*, *republican*, and *likely*. The adverb *likely* is erroneously classified as an adjective, and the words *hot* and *rich* are treated as relating strongly to smell, presumably because they frequently occur in food-related contexts. Even more counterintuitive classifications characterize the auditory modality: The ten "adjectives" classified as most auditory are *derisory*, *assisted-suicide*, *non-proliferation*, *dmcneely*, *fast-selling*, *fractional*, *guffawing*, *held*, *b.s.*, and *litton-made*. Not only are many of these words not adjectives, but their status as being auditory is highly questionable. In general, besides dubitable sensory classifications, the Sensicon also includes a lot of inconsistent part-of-speech tags, such as *domestic-sales*, *ski-slope*, *two-out*, and *serial-killer* being treated as adjectives.

Compared to the Sensicon classifications, the Lynott and Connell (2009) norms can be regarded as a human gold standard. Of the 423 sensory adjectives from Lynott and Connell (2009), 412 are also contained in the Sensicon. Of these, only 161 (39%) have the same modality classification. Conversely, this means that 251 (61%) have mismatching dominant modality classifications. In particular, the Sensicon classifies many more words as gustatory. For instance, the words *rectangular*, *cloudy*, and *ugly* are classified as dominantly visual in the Lynott and Connell (2009) norms but as gustatory in the Sensicon. Pairwise correlations between the Sensicon and the modality norms are really low, with Pearson's r = 0.37 for sight, r = 0.27 for touch, r = 0.32 for sound, r = 0.51 for taste, and r = 0.58 for smell.

Is it at all appropriate to use co-occurrence information for sensory classification? Chapter 16 shows that taste and smell language specializes in evaluative functions. As a result, words may be classified as gustatory purely because they are evaluative. In fact, the word *bad* is classified as gustatory in the Sensicon. Moreover, taste and smell words are associated with each other, as are sight and touch words (Louwerse & Connell, 2011; Lynott & Connell, 2009; see also Chapters 13–16). This means that any approach purely based on text co-occurrences will have difficulty distinguishing between these two modality pairs. Co-occurrence based approaches have further problems disentangling metaphor and literal uses of sensory words, such as when using the touch word *sharp* to talk about sounds (Tekiroğlu, Özbal, Strapparava, 2015). Moreover, associations between sensory words are contingent on text type, as shown by Strik Lievers (2015). This would suggest that different corpus choices will lead to different classifications. In general, any co-occurrence based method is highly susceptible to the data from which co-occurrence statistics are computed.[4]

One may think that the human ratings in Lynott and Connell (2009) are more subjective than the seemingly objective automatic technique used to construct the Sensicon. However, it needs to be pointed out that the Sensicon is still based on subjective judgments—namely, based on the introspection of the lexicographers who compiled FrameNet and WordNet, which were used to construct the seed list based on which co-occurrence statistics were computed. Given this, one could argue that the Lynott and Connell (2009) norms are a more direct and transparent reflection of intuitions. Furthermore, the Sensicon norms cannot be used for corpus analyses precisely because the norms are corpus-based, which would lead to circularity.

Finally, compared to any other word list, the existing modality norm set by Lynott and Connell (2009) has the advantage that their utility has been validated in several psycholinguistic experiments, including Connell and Lynott (2010, 2012, 2014, 2016). For example, Connell and Lynott (2012) showed that the maximum perceptual strength value of the norms is a better predictor of word processing times than comparable concreteness and imageability ratings. Finally, Connell and Lynott (2011) showed a modality switching cost (Pecher et al., 2003) in a concept creation task using words classified according to the norms considered here. Connell and Lynott (2014) also used the modality norms to show that lexical decision and reading-aloud tasks direct attention to vision, but reading-out tasks additionally direct attention to audition. For additional experimental demonstrations of other norm sets that are structurally equivalent to the presently considered norms, see van Dantzig et al. (2011), Connell and Lynott (2016), and Speed and Majid (2017). For word lists such as those by Strik Lievers (2015) or the Sensicon, no such empirical demonstrations exist.

To conclude, there is a wealth of advantages to relying on native speaker ratings and, in particular, the dataset of 423 adjectives from Lynott and Connell (2009). This dataset, in addition to other datasets of modality norms, is an important component of the methodological toolkit of sensory linguistics. Its many uses will be demonstrated throughout the remainder of this book.

---

[1] In fact, Medler et al. (2005) gathered the first modality norms that I am aware of, but they only normed the dimensions of sound and color, among other semantic dimensions that do not have anything to do with the five senses.

[2] The rating responses were not normalized within participants. When asked to rate something, participants often use the rating scale differently; for example, while some tend to use the whole scale (from 0 to 5), others selectively focus on the central part of the scale (e.g., from 1 to 4). Therefore, when averaging the ratings between participants, some information might be lost. It

would be interesting to see how the results presented here compare to a normalized scale and whether there are systematic individual differences between participants that relate to sensory linguistics.

[3] Similar to the task in Lynott and Connell (2009), participants in the study by van Dantzig et al. (2011) had to rate each conceptual pair for each of the five sensory modalities. This yields a data structure that mirrors the one of Lynott and Connell (2009), with visual strength ratings, auditory strength ratings, et cetera—except each adjective now has two such ratings, one for each noun it was combined with in the stimulus set (e.g., *abrasive lava, abrasive sandpaper*). For the van Dantzig et al. (2011) norms, the average of the responses for each of the two contexts was computed. In the case of the tactile modality and *abrasive* this would be 3.59, based on the mean of *abrasive sandpaper* (4.81) and *abrasive lava* (2.37).

[4] The amount of data going into the computation for each word is another issue to consider as well. Although the classifications in the Sensicon are based on a normalized form of pointwise mutual information, they are still susceptible to how much data is available for each word. A corpus gives more information about co-occurrence statistics for high frequency words than for low frequency words. It is possible this may actually impact the results; for sight, the correlation between the Sensicon values and the modality norms from Lynott and Connell (2009) is lower for low frequency words (using a median split and word frequency data from SUBTLEX, sight HF: r = 0.40, LF: r = 0.30). For taste (HF: r = 0.38, LF: r = 0.66) and smell (HF: r = 0.43, LF: r = 0.67), the pattern is the reverse. Here, low frequency words in the Sensicon have much higher correlation with human ratings. Touch (HF: r = 0.29; LF: r = 0.23) and sound (HF: r = 0.34; LF: r = 0.36) are inbetween.

Chapter 12. Dominance relations and specialization of sensory words

12.1. Introduction

This chapter delves more deeply into the modality norms by Lynott and Connell (2009). The first part of this chapter focuses on overall dominance relations between the senses (Chapter 12.2). The second part focuses on the degree to which sensory words are multisensory (Chapter 12.3).

Several of the analyses presented here and in the following chapter recapitulate analyses already conducted by Lynott and Connell (2009, 2013), as well as by Winter et al. (2018) and Strik Lievers and Winter (2018). Repeating some of these analyses serves to familiarize the reader with the 423 adjectives that will be used throughout the remainder of the book. These chapters also go beyond existing work by introducing new analyses, such as analysis of the precise distributional characteristics of the different perceptual strength rating measures. Moreover, I will reconceptualize some of the work that Lynott and Connell (2009, 2013) have conducted on these norms.

Two theoretical topics will be investigated: First, the ratings will be investigated with respect to visual dominance and smell inferiority (see Chapter 4). Second, the idea that sensory words are compression devices that single out modality-specific perceptual content will be tested with the rating data (see Chapter 4).

12.2. Dominance relations between the senses

12.2.1. Predictions

As was discussed in Chapter 4, a lot of research suggests that vision is the dominant sense not only in perception, but also in human language, and the opposite is true for smell (Levinson & Majid, 2014). Therefore, it is expected that the ratings from Lynott and Connell (2009) exhibit visual dominance as well as smell inferiority. The other senses are expected to pattern in between. Crucially, because the rating dataset is a rich multidimensional dataset, dominance relations between the senses may surface in a variety of ways. The following aspects of visual dominance (and smell inferiority) will be explored in this chapter:

(a) visual dominance in continuous perceptual strength ratings (Chapter 12.2.2)
(b) visual dominance in categorical word counts (word types) (Chapter 12.2.3)
(c) visual dominance in the distributional characteristics of the perceptual strength ratings (Chapter 12.2.4)

12.2.2. Dominance in perceptual strength ratings

When looking at the continuous perceptual strength ratings, the following ranking emerges: Sight received the highest ratings overall (*M* = 3.6, *SD* = 1.2), followed by touch (*M* = 2.2, *SD* = 1.6). By comparison, sound (*M* = 1.5, *SD* = 1.6), taste (*M* = 1.3, *SD* = 1.6) and smell (*M* = 1.2, *SD* = 1.5) had much lower ratings. These means are depicted in Figure 3.

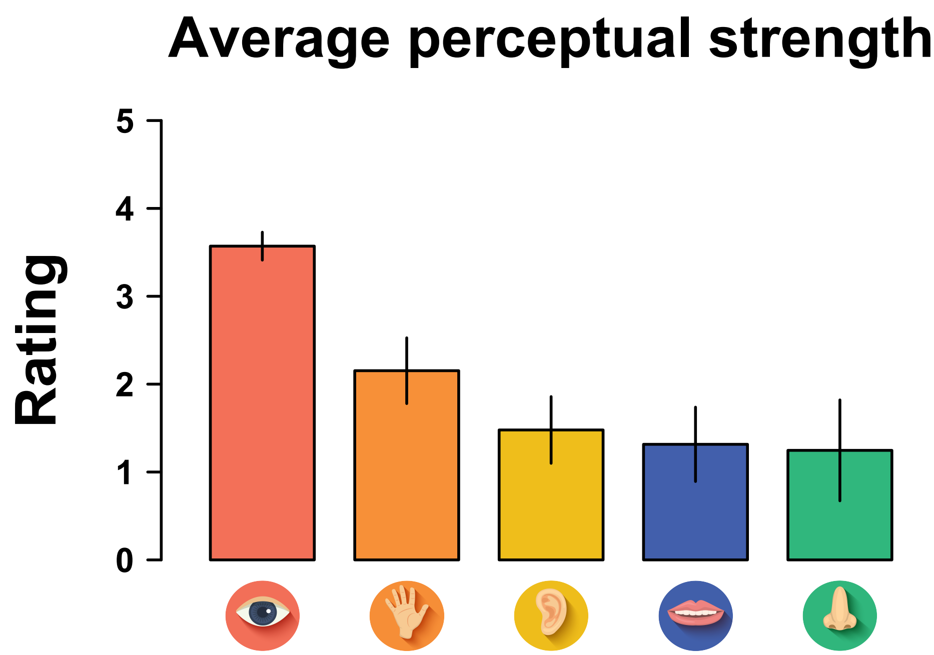


Figure 3. Average perceptual strength ratings per modality from Lynott and Connell (2009). Error bars indicate 95% confidence intervals of the respective means. For more detailed distributional information, see Figure 4.

Are these perceptual strength ratings reliably different from each other? To assess this, a series of paired Wilcoxon signed rank tests was used, one for each modality pair (such as sight versus smell, sight versus taste, etc.).[1] This showed that sight had reliably higher ratings than any of the other modalities (all Dunn-Šidák corrected p's < 0.05). Moreover, touch had reliably higher ratings than sound, taste, and smell. Taste and smell were not reliably different from each other ($V = 26106$, corrected $p = 0.11$), and neither were sound and smell ($V = 47900$, corrected $p = 0.12$), or sound and taste ($V = 46980$, corrected $p = 0.57$). These results suggest a split between sight and touch on the one hand, and taste, smell, and sound on the other.

12.2.3. Dominance relations in categorical word counts

The categorical modality classifications, which are based on each word's maximum perceptual strength rating (see Chapter 11), can be used to further explore dominance relations between the senses. This allows us to count how many words there are for each sense and whether some senses have more words than others. The analyses of categorical word counts presented in Winter et al. (2018) and Strik Lievers and Winter (2018) will be repeated here for the sake of exposition. It should be borne in mind, however, that these discrete word counts ignore the fact that sensory words are multisensory.

Using the dominant modality classifications, there are 205 sight adjectives, 70 touch adjectives, 68 sound adjectives, 54 taste adjectives, and 26 smell adjectives (see Lynott & Connell, 2009). A simple Chi-Square test indicated these counts to be reliably different from each other ($\chi^2(4) = 228.8$, $p < 0.0001$). Adjusted standardized Pearson residuals can be used to assess which senses were reliably overrepresented, and which were reliably underrepresented (using $|2|$ as a cut-off for significance, see Levshina, 2015, pp. 220–221). Sight was reliably overrepresented (+14.6). Smell was the most underrepresented (–7.1), followed by taste (–3.7). Sound was reliably underrepresented as well (–2.0), and touch was found to be neither reliably overrepresented nor reliably underrepresented (–1.8).

12.2.4. Dominance in distributional characteristics

So far, these analyses looked at the overall perceptual strength ratings, either using the continuous ratings (Chapter 12.2.2) or treating the ratings in a categorical fashion (Chapter 12.2.3). However, dominance relations may also express themselves in the characteristics of the perceptual strength rating distributions. In particular, it has to be borne in mind that participants had a continuous range available to classify sensory words, ranging from 0 (no modality association) to 5 (maximal modality association). Because the scale is continuous, participants are free to use any one part of the range. It could be that participants' responses are clustered around certain values, and this degree of clustering may differ between the five senses.

Figure 4 shows the distributions of the perceptual strength ratings for each modality. The *x*-axis corresponds to the perceptual strength ratings (from 0 to 5). The *y*-axis corresponds to the density, reflecting how many words there are for a particular value of the scale. The solid vertical lines indicate the means of each distribution.

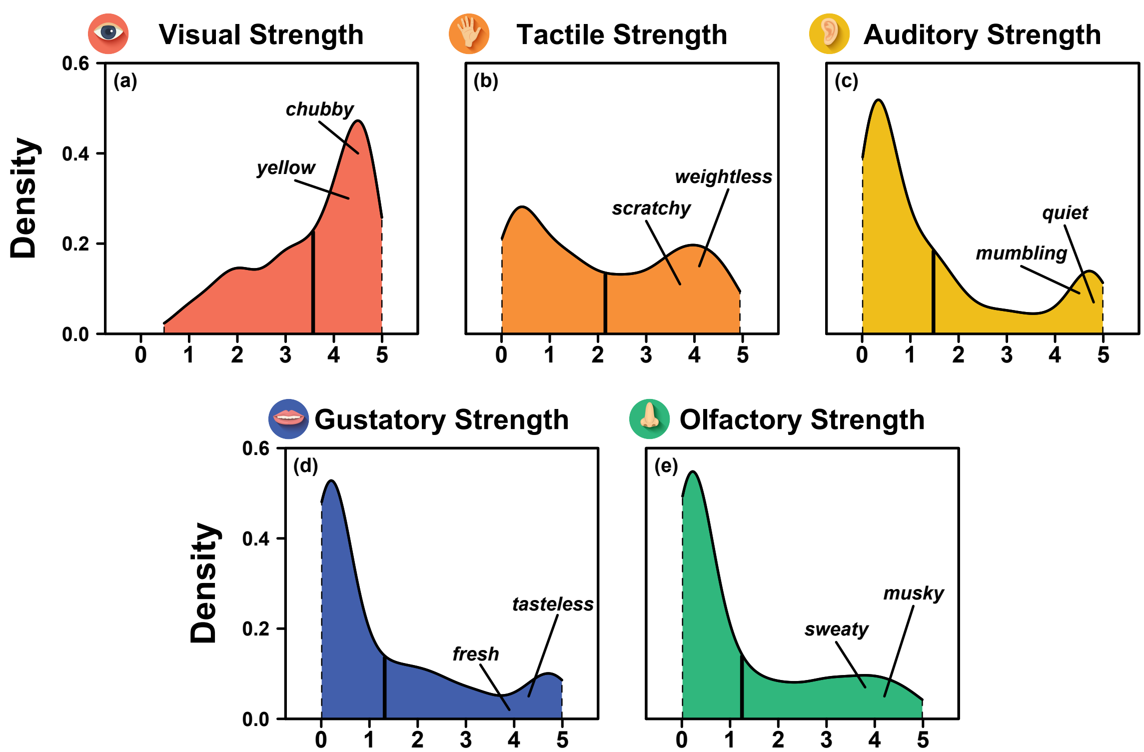


Figure 4. Kernel density estimates of perceptual strength ratings for the Lynott and Connell (2009) adjectives. Density curves are restricted to the observed range; solid vertical lines indicate means. The curves show clear bimodality (two-peakedness) for all senses except for sight.

In part, Figure 4 shows what was also seen in the average perceptual strength ratings, with overall higher sight ratings than ratings for the other modalities (see solid vertical lines). However, Figure 4 also shows that the perceptual strength rating distributions differ in *shape* between the five senses. In particular, touch, sound, taste, and smell appear to be two-peaked (i.e., they are bimodal distributions). This says something meaningful about the sensory vocabulary of English: For all non-visual modalities, there are dedicated pockets of lexical material. For example, the words *quiet* and *mumbling* strongly relate to sound, and they form a cluster of sound words together with words such as *groaning*, *thunderous*, and *purring*. However, there are many more words that do not relate to sound at all, and these form a separate cluster with its own peak in the distribution visualized in Figure 4c. Sight words are not similarly restricted to small pockets of lexical material, which is suggested by the fact that the visual strength ratings have only one peak. The lack of a two-peaked distribution for visual strength ratings suggests that all words in the study of Lynott and Connell (2009) are somewhat visual. This can be interpreted to show that vision is less restricted within the sensory vocabulary of English.

Hartigan's dip test (Hartigan & Hartigan, 1985) revealed statistically reliable bimodality for touch (D = 0.03, corrected p = 0.004), sound (D = 0.04, p = 0.0001), taste (D = 0.04, p = 0.0009), and smell (D = 0.04, p = 0.0009), but not for sight (D = 0.02, p = 0.99).[2] This finding can be interpreted as visual dominance in the distributional characteristics of perceptual strength ratings: All non-visual modalities are restricted to small dedicated portions of the sensory lexicon.

Taking all results on dominance relations together, Chapter 12.2 showed visual dominance in the form of overall perceptual strength ratings (Chapter 12.2.2), word type counts (Chapter 12.2.3), and rating distributions (Chapter 12.2.4). On the other hand, this chapter has also shown evidence for the linguistic inferiority of taste and smell in English. In addition, the results suggest an asymmetry between touch and sound, with touch ratings higher than sound ratings and with more touch than sound words. The sound ratings also exhibited more bimodality than the touch ratings, which can be taken to suggest that sound is more restricted to a small pocket of highly exclusive lexical material. This composition of the sensory vocabulary has to be taken into account, for example, when performing studies of crossmodal language use (see Chapter 17).

12.3. Modality exclusivity

12.3.1. Specialization of sensory vocabulary

The last section discussed dominance relations between the senses. This section looks at the degree to which words specialize into any one sensory modality. As was discussed in Chapter 4, sensory words can be seen as compression devices that single out particular aspects of our perceptual worlds. On the other hand, I have also argued that perception is intensely multisensory. In their original rating study, Lynott and Connell (2009) highlight that property words are multisensory, citing the fact that the average modality exclusivity of all words was 46% as evidence for this claim.

However, it is possible that the figure 46% is mathematically inevitable. It could be that randomly picking perceptual strength ratings would result in an average exclusivity close to 46%. One has to consider the fact that noise in the ratings would show up spuriously as multisensoriality. For example, if participants did not know a word very well, they might merely guess when performing the rating task. This is going to lead to the appearance of multisensoriality because on average, the ratings will be about equal for all modalities. In general, what is a high or low modality exclusivity to begin with? That is, what acts as an appropriate baseline for this figure and hence, for assessing the degree of multisensoriality in property words?

12.3.2. A baseline for modality exclusivity

One way to construct a meaningful baseline is to compute the modality exclusivity of the average word. This hypothetical average word is what was visualized in Figure 3. Mathematically, this word is represented by the modality vector [3.6, 2.2, 1.5, 1.3, 1.2], with average perceptual strength values in respective order of sight, touch, sound, taste, and smell. The modality exclusivity of this hypothetical word is 24% (range divided by sum times 100).[3] This number can be used as a hypothetical baseline against which to compare the observed exclusivity values from the Lynott and Connell (2009) dataset. A one-sample Wilcoxon signed rank test showed that this was indeed the case ($V = 88342$, $p < 0.0001$).

In substantive terms, this means the following: Although adjectives are clearly multisensory, they are actually *more* exclusive than is expected by chance alone. This is exactly what we would expect if sensory adjectives preferentially express content from a particular sensory modality. Lynott and Connell (2009) rightly pointed out that adjectives are rarely ever specialized in *just* one sense. However, the degree of multisensoriality is not unlimited, and the exclusivity ratings do suggest some specialization of sensory adjectives.

12.3.3. A better baseline for modality exclusivity

The results presented in the last section are already suggestive of specialization; however, the statistical baseline used was not optimal. In particular, the averages represented by the vector [3.6, 2.2, 1.5, 1.3, 1.2] ignore the distributional characteristics visualized in Figure 4. A more adequate baseline would compare the observed exclusivity values against hypothetical exclusivity values that obey the distributional characteristics of the data. This can be done via using a permutation-based approach.

In general, permutation-based approaches work by reshuffling a dataset to create a statistical baseline. In this case, I shuffled the lists of perceptual strength ratings separately for each modality. This means that each word has a random set of perceptual strength values. However, all the original perceptual strength values are still represented within the dataset, just not paired with the correct words. Then, the exclusivity was computed for each word. Following this, the average exclusivity was computed for all of these random words. The result is a single modality exclusivity value for a randomly generated sensory vocabulary. In this reshuffled data, each perceptual strength measure has the same distribution as seen in Figure 4, but the perceptual strength values do not match up between the senses the same way they do in the real dataset. This procedure is then repeated 1,000 times, each time storing the chance-based modality exclusivity of each hypothetical sensory vocabulary. This generates a distribution of hypothetical modality exclusivity values against which we can compare the average exclusivity value that is actually observed in the adjective dataset—that is, the figure 46%.

Figure 5a plots the distributions of the permutation-based modality exclusivities, based on repeatedly shuffling the adjective data. For comparison, Figure 5b shows the same approach used on the noun data from Lynott and

Connell (2013). The dashed vertical lines are positioned at the observed average exclusivity values for each dataset.

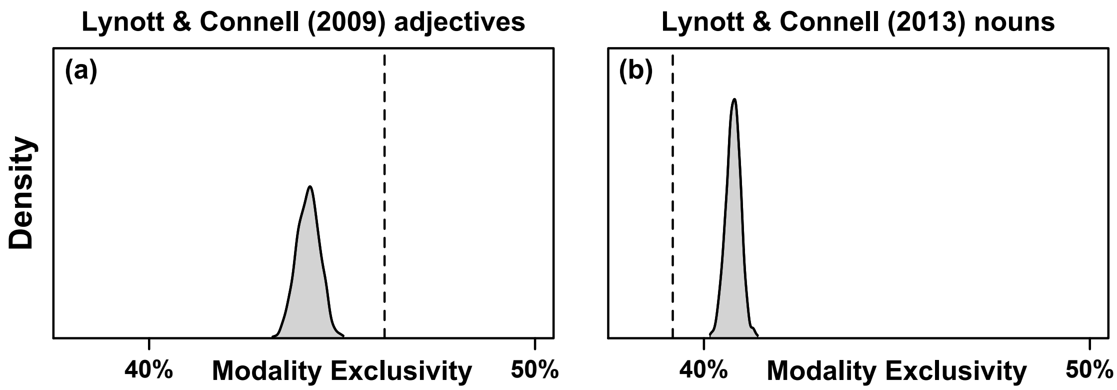


Figure 5. Kernel density estimates of randomly permuted exclusivities. Vertical dashed lines indicate the observed exclusivity for each dataset.

Figure 5a shows that the average exclusivity value of 46% is higher than all of the permuted exclusivities. The average of all permuted exclusivities was 44%, which is more multisensory than the actually observed exclusivity. In this case, the permutation-based p-value is exactly 0.00; that is, there is no single permuted exclusivity that is higher than the actually observed exclusivity. This can be interpreted to mean the following: Sensory adjectives are reliably *less* multisensory than what is expected based on a chance baseline; that is, they are unexpectedly exclusive. In contrast, Figure 5b shows that the average exclusivity of the nouns (39%) is lower than all baseline exclusivity values, which has a mean of 41% (p = 0.00). Thus, this analysis indicates nouns to be reliably *more* multisensory than what is expected by chance alone. These results provide further evidence for the idea that sensory adjectives specialize in particular modalities.

This analysis thus suggests that the figure 46%—the average exclusivity of all sensory adjectives taken together—is potentially deceiving. While Lynott and Connell (2009) interpret this figure to indicate that sensory adjectives are multisensory, the present analyses suggest that despite this multisensoriality, there is a drive toward specialization, or what was under the banner of "compression" in Chapter 4.

12.3.4. Modality exclusivity differences between the senses

Lynott and Connell (2009) already noted that the senses differ in the degree to which they are exclusive or not. When splitting up the data by dominant modality, sound words have overall higher modality exclusivities (57%) than sight words (49%), followed by smell words (42%), touch words (37%), and taste words (35%). A Kruskal-Wallis test indicates that these average exclusivities are reliably different from each other ($\chi^2(4) = 97.09$, $p < 0.0001$).

The permutation-based approach can also be used to see whether there is evidence for an emphasis on specialization or multisensoriality within each set of sensory words. To do this, the permutation-based approach was repeated for subsets of words based on their dominant modality (i.e., separately for the 205 sight adjectives, for the 70 touch adjectives, the 68 sound adjectives etc.). The results are shown in Figure 6.

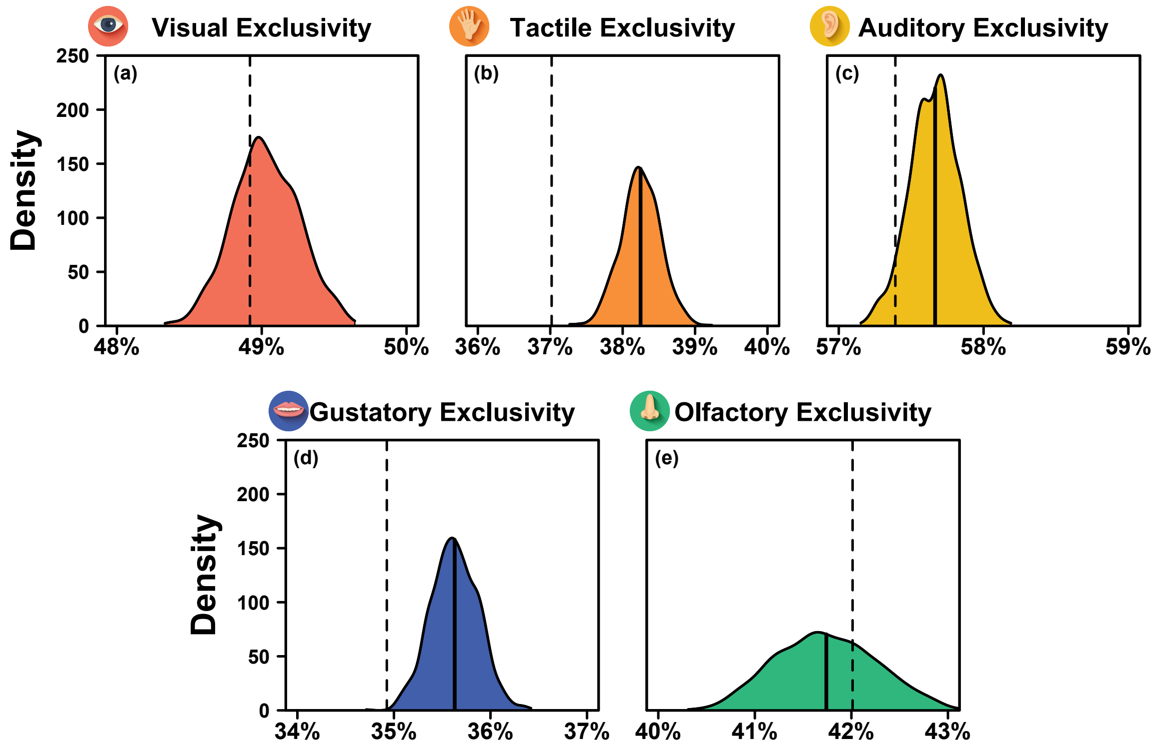


Figure 6. Kernel density estimates of randomly permuted exclusivities per modality. Vertical dashed lines indicate the average exclusivity of the observed data for each sense.

To interpret Figure 6, one needs to compare the position of the permuted distributions to the dashed lines, which indicate the average exclusivity values for each sensory modality that are actually observed in the data. As can be seen, observed averages were very similar to the permutated values for sight, sound, and smell. For these sensory modalities, there was no evidence that the observed exclusivity differs from chance expectation. Only for touch (permutation-based p-value = 0.0) and taste (p = 0.02) did the permuted distributions differ markedly from the observed average exclusivity. In both cases, the permuted distributions had *higher* exclusivities than the observed average (i.e., random words were less multisensory). This indicates that the actually observed exclusivities of touch and taste words are indeed unexpectedly low (i.e., specifically these two modalities are highly multisensory).

12.4. Conclusions

I have presented two sets of results in this chapter. The first set of results showed not all senses are created equal. The perceptual strength ratings exhibited evidence of visual dominance, as well as clear evidence for smell ineffability. Taste and sound words also showed somewhat weaker signs of ineffability.

We can take the word type counts of dominant modality classifications as a measure of lexical differentiation. The type counts corroborated the view that there is visual dominance in the English lexicon (Buck, 1949; Levinson & Majid, 2014; Viberg, 1983; Winter et al., 2018). Having the most unique word types, sight is the most differentiated sense in English. This also means that speakers can make more fine-grained semantic distinctions in the visual modality compared to the other modalities. It should furthermore be noted that the chemical senses were the most underrepresented in type counts, particularly smell. The distributional analysis (see Figure 4) additionally reflected visual dominance: Whereas the other senses are confined to small pockets of lexical material, there is no such restriction for sight.

It is noteworthy that sound was not particularly dominant in the present analyses. For example, an analysis of the perceptual strength measures failed to find a statistically reliable difference between the perceptual strength ratings for sound and the perceptual strength ratings for smell. That is, sound did not have higher ratings than smell, a modality that is generally thought to be one of the "lower" senses. This seems to stand against the idea, often expressed in research on synesthetic metaphors (see Chapter 8), that sound is a "higher" sensory modality. We also saw that sound vocabulary is the most restricted in the Lynott and Connell (2009) dataset. This is not only evidenced by high modality exclusivity values, but also by the fact that sound ratings exhibited the most pronounced bimodality, with a small pocket of dedicated sound words against a backdrop of many words devoted to the other senses.

Another contribution of this chapter was to quantify and contextualize the idea that sensory words are multisensory. This was achieved by investigating the modality exclusivity measure, which was first introduced by Lynott and Connell (2009). A re-analysis of their measure showed that even though sensory adjectives are multisensory, they are actually *more* exclusive than what is expected by chance. This indicates specialization into modality-specific perceptual content. Sensory adjectives carve up the perceptual world into smaller, more manageable chunks. This means that each sensory word tends to focus on one sensory modality, or a particular set of perceptual qualities, at the expense of the other modalities or other perceptual qualities (Chapter 4). This finding, together with the evidence for multisensoriality, suggests that sensory adjectives are neither fully multisensory nor fully specialized. They can be seen as occupying a sweet spot between these two extremes.

In further analyses, I looked at modality exclusivity separately for each sensory modality. As was already reported in Lynott and Connell (2009), touch and taste emerged as the most multisensory domains. Sensory words associated with these modalities had the lowest average modality exclusivity. This finding corresponds to the observation that taste and touch words are particularly prone to being used crossmodally (Chapter 8).

The next chapter looks at inter-relations between the perceptual strength ratings. How do the senses work together in this set of sensory adjectives?

[1] Wilcoxon tests were used with continuity correction; *p*-values are Dunn-Šidák corrected for performing multiple comparisons (10 tests for each modality pair).

[2] Because Hartigan's dip test was applied five times to five different datasets, all *p*-values were corrected for five tests (Dunn-Šidák correction).

[3] Given the observed means in the dataset, this exclusivity is actually the lowest *average* exclusivity that could possibly be observed. Particular individual words may be below this value (such as the word *strange* [9.6%] or *harsh* [11.6%]), but it is unexpected that the average exclusivity across all words would go below 24%.

Chapter 13. Correlations and clusters

13.1. Introduction

This chapter continues the analysis of the modality norms provided by Lynott and Connell (2009). Chapter 11 discussed how native speakers had to rate each word on each of the five senses separately. This would suggest that the 423 adjectives are structured in terms of the five senses folk model. However, is it actually the case that a description of this dataset in terms of five senses is the best we can do? Moreover, does the dataset actually provide evidence for five subgroups of words that are neatly packaged into the five senses? Or is there statistical evidence for smaller or larger groupings?

This chapter looks at correlations and clusters within the modality norms. These analyses speak directly to the question of whether the five senses folk model can be applied straightforwardly to the sensory vocabulary of English (Chapter 2). The chapter is structured as follows:

(a) correlations, uncovering larger subgroups (Chapter 13.2)
(b) clusters, uncovering smaller subgroups (Chapter 13.3)

13.2. Correlations between the senses

Lynott and Connell (2009) used Principal Component Analysis (PCA) on the matrix of perceptual strength ratings to look at interrelations between the senses. PCA is a dimensionality reduction technique. Dimensionality reduction is best explained in analogy with the night sky: Physical space has three dimensions, but the night sky presents itself to us as a two-dimensional plane. This means that three-dimensional space is projected (reduced) onto two dimensions. Celestial bodies that are millions of lightyears away from each other along the third dimension, the dimension of depth, appear next to each other on the night sky. Nevertheless, the night sky does capture some of the major relations between celestial bodies that are of interest to us as humans. Even just two dimensions give us considerable information about the relative positioning of certain galaxies and stars.

Similar to projecting three-dimensional physical space onto a two-dimensional night sky, PCA projects a multidimensional dataset into a lower-dimensional space. Elements of the data that reliably vary together will be projected onto the same dimensions, called components. In the present case, this can be used to project the "space" spanned by the five modality vectors (one for each sense) onto a lower-dimensional space. Essentially, this is asking the question: Is it possible to capture the major relations between sensory words with less than five senses?

To answer this question, a PCA (using singular value decomposition) was run on a matrix of the *z*-scored perceptual strength ratings.[1] Table 7 shows how each of the five senses loads onto each component.

Table 7

Principal components, loadings, and variance explained

Modality	**C1**	**C2**	**C3**	**C4**	**C5**
Sight	+0.3	–0.6	–0.5	+0.6	0.0
Touch	+0.2	–0.5	+0.8	0.0	–0.2
Sound	+0.3	+0.6	+0.3	+0.7	–0.1
Taste	–0.6	–0.1	+0.2	+0.4	+0.7
Smell	–0.6	–0.1	0.0	+0.2	–0.7
Variance	42%	75%	88%	96%	100%

To interpret this table, one needs to look at the sign and the strength of each loading. This reveals that the first component distinguishes sight, touch, and sound from taste and smell. This component accounts for 42% of the variance in ratings. The second component distinguishes sight and touch from sound (taste and smell do not load strongly onto this component). This component accounts for an additional 33% off the variance in ratings. Together, the first two components account for 75%.

The third component separates sight from touch (13%). The fourth component has high loadings for sight and sound (8%). Finally, the fifth component separates taste and smell (4%). It is interesting to note that the separation of taste and smell words from each other describes the least variance, whereas the separation of taste and smell words from the rest describes much more variance. This suggests that the chemical senses of taste and smell are strongly associated with each other in the sensory vocabulary; separating the two modalities from each other does not help as much in describing this dataset. Similarly, the PCA indicates touch and sight ratings to be correlated with each other, with the distinction between these two modalities being comparatively less important (they have strongly different loadings only for the third component).

Figure 7 shows the variable correlation plot for the first two components. In this plot, each sensory modality is represented as a vector. If two vectors have similar angles (i.e., pointing in the same direction), the corresponding ratings are highly correlated with each other. By moving from left to right along the x-axis, we move from words that are both taste and smell to words that are sight, touch, and sound-related. Moving along the y-axis (from top to bottom), we differentiate sound from touch and sight. Taste and smell do not load heavily onto the second component. This plot makes it immediately apparent that sound points in a direction away from the other sensory modalities, which speaks to the exclusivity of this sensory modality.

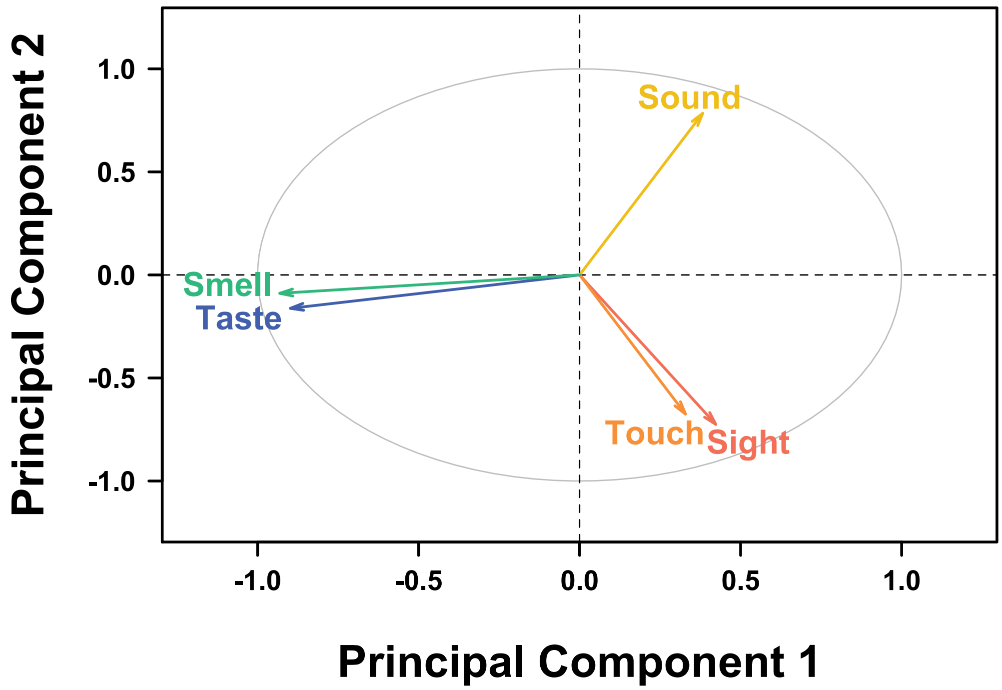


Figure 7. Variable coordinate plot for the first two components of the Principal Components Analysis. Vectors pointing into the same direction indicate variables that are highly correlated with each other.

Figure 8 shows a scatter plot of all sensory words within the two-dimensional space spanned by the two components. Each data point represents a single word, colored according to the dominant modality classification. This visualization clearly shows that in the two-dimensional projection, taste and smell, as well as touch and sight, are not clearly distinguished. Most of the taste words and smell words belong to what visually presents itself as just one cluster;

the same applies to sight and touch. The logos representing each sense (nose, mouth, etc.) are located at the *x,y* means of the respective dominant modalities.

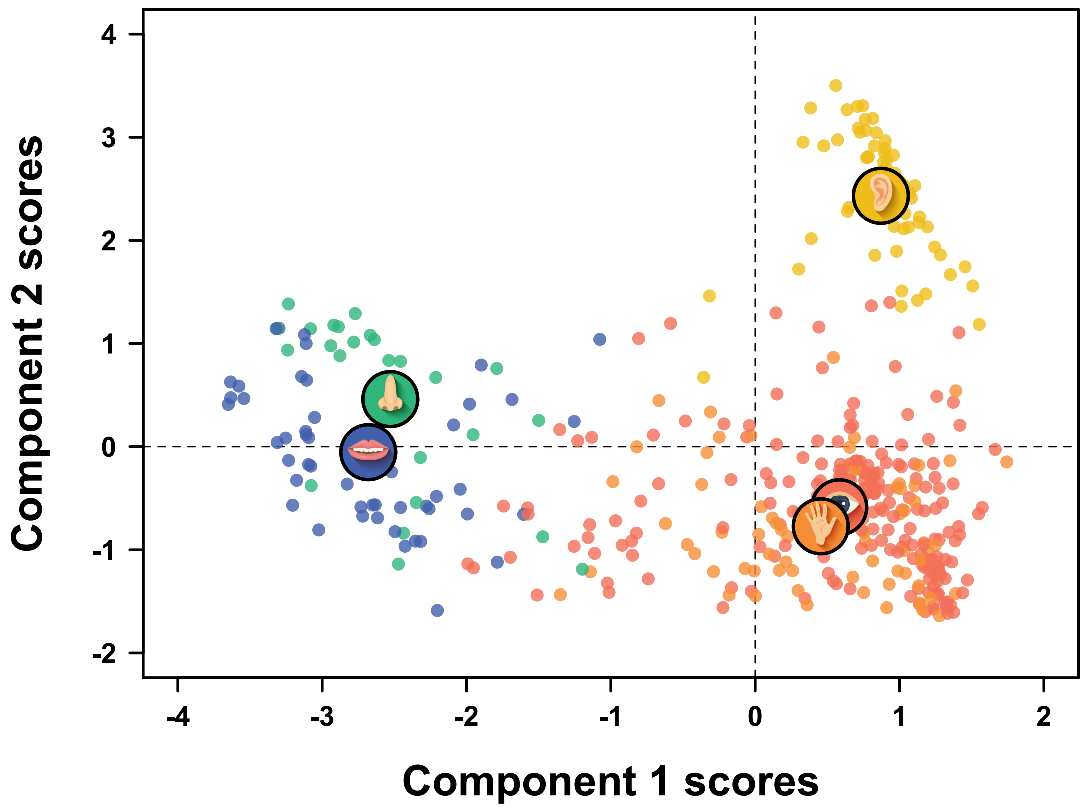


Figure 8. Scatter plots of words within the two-dimensional space spanned by the first two principal components. Each data point represents a word; sense logos are located at the means of each dominant modality.

The PCA furthermore suggests that sound is quite distinct from the other senses. Its correlation vector in Figure 7 points away from all other sensory modalities. This was also observed by Lynott and Connell (2009) and replicated by Lynott and Connell (2013): Sound concepts appear to be quite exclusive; they are heavily tied to their own modality. This pattern will be revisited in Chapter 14.

This section looked at meaningful macrostructures within the modality space spanned by the five senses. Correlations between the senses helped us to uncover larger groupings. The resulting clusters crisscross the five senses folk model: Taste and smell belong together, and so do touch and sight. Sound appears to form a group by itself. In the next section of this chapter, we move from uncovering larger subgroups to uncovering smaller subgroups. In total, this set of analysis looks at the rich multidimensional data provided by Lynott and Connell (2009) from multiple different lenses to uncover hidden structures at various levels of granularity.

13.3. Clustering the senses

Applying a cluster algorithm to the perceptual strength ratings allows the detection of latent classes (groups of words that have similar perceptual strength ratings). Cluster analysis attempts to find groups of data points that have similar numerical characteristics. There are several different methods available for cluster analysis. Here, Gaussian mixture models (GMMs) will be used.[2] A series of such models was run on the matrix of perceptual strength ratings (*z*-scored). Bayesian Information Criterion scores (BICs) were used to assess which number of clusters provided the optimal fit to the data. It turned out that the data best supported a 12-cluster solution.[3] This means that there is statistical evidence for 12 distinct subgroups that have highly related perceptual strength ratings.

In the remainder of this section, I will go through the clusters one by one. In doing so, it has to be borne in mind that each cluster is just a statistical construct—a set of words that hang together based on similar 'profiles' of their perceptual strength ratings. Whether the words lumped together form meaningful groupings can only be determined later by means of human interpretation. To keep track of the twelve clusters, I labelled them in a heuristic fashion, using verbal labels that serve to summarize the essential characteristics of each cluster.

Figure 9 shows the first two clusters. The barplots show average perceptual strength ratings, which can be interpreted as a snapshot of the modality profile of each cluster. The examples below each barplot list the ten adjectives that were most strongly associated with each cluster (they are ranked according to the certainty of cluster association).

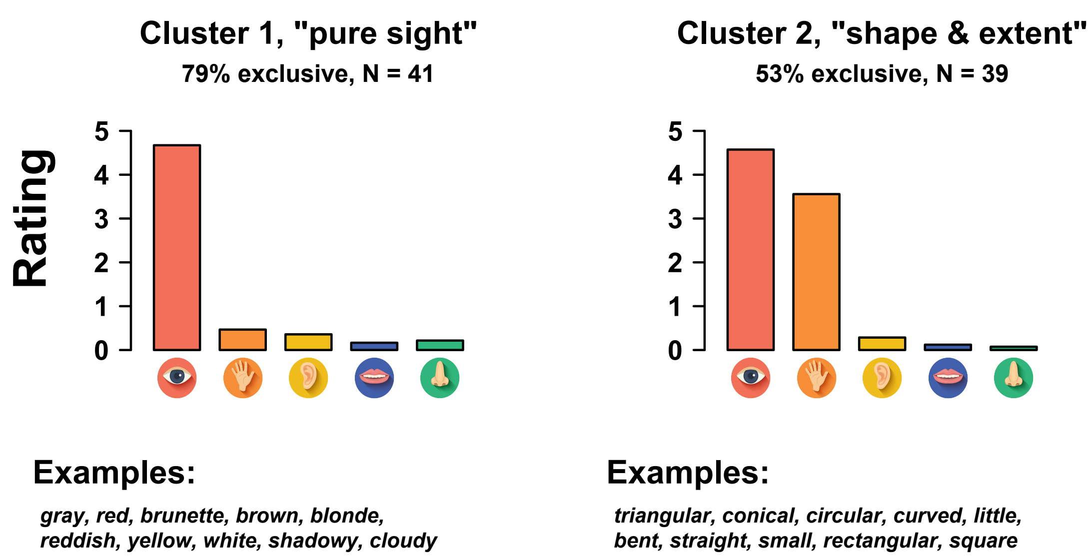


Figure 9. Clusters 1 and 2. from the Gaussian mixture model**.** Words are ranked by the degree of certainty with which they associated with each cluster (i.e., the words that are most certainly a member of this cluster are mentioned first).

The "pure sight" cluster contains 41 words. The bar plot of the perceptual strength ratings clearly shows that the words in this cluster are similar to each other by virtue of having very high visual strength ratings and low ratings for all of the other senses. Given this specialization into sight, it also comes as no surprise that the words in this cluster are highly exclusive (79%). Among the pure sight words, there are color terms, such as *gray*, *red*, and *brown*, and words that indicate brightness or visibility conditions, such as *shadowy* and *cloudy*.

The "shape & extent" cluster contains 39 words that have overall less exclusivity (53%). These words appear to refer to shapes (*triangular*, *conical*, *circular*), edginess (*curved*, *bent*, *straight*) or extent (*little*, *small*). The words in this cluster are reminiscent of the dimension-word category by Williams (1976; Chapter 4.3.2; Chapter 8.2). These words describe the gross dimensions of objects in terms of shape and size. Since any perceptual characteristic that directly relates to physical space is a common sensible for sight and touch (see Marks, 1978, Chapter 2; Stokes & Biggs, 2015), the "shape & extent" cluster has high ratings for both sight and touch. Louwerse and Connell (2011, p. 384) say that "any object that can be touched can be seen." The fact that shape and extent are common sensibles also explains the relatively low exclusivity of this cluster (53%). Interestingly, the "shape & extent" cluster has overall higher sight than touch ratings. This suggests that to native speakers of English at least, spatial words first and foremost appear to be visual.

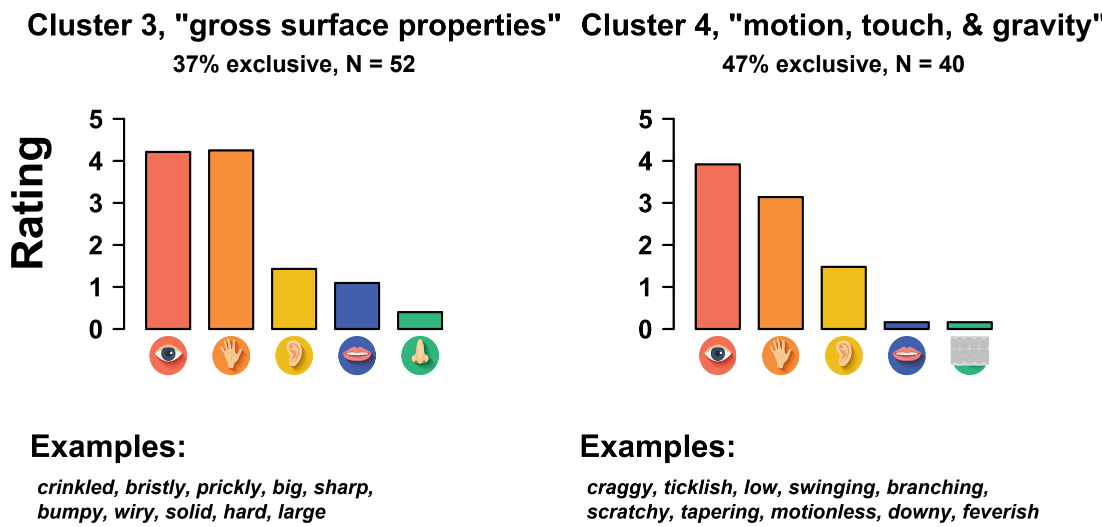


Figure 10. Clusters 3 and 4. Words are ranked by their certainty

Figure 10 shows the next two clusters. The "gross surface properties" cluster contains 52 words. The "motion, touch, & gravity" cluster contains 40 words. With average exclusivities of 37% and 47%, both clusters are considerably more multisensory than the other two clusters discussed so far. The descriptors in cluster 3 are all labels that would easily apply to things in a garage or a toolshed. These words describe the general shape and texture of physical objects. These terms are similar to the "shape & extent" cluster, but they appear to involve more words that involve irregular surfaces (*crinkled*, *bristly*, *prickly*, *bumpy*, *wiry*), and their focus seems to be overall more on surface properties (*solid*, *hard*), with comparatively fewer genuine size words (*large*, *big*). The words in the "motion, touch, & gravity" cluster can be characterized along various dimensions. They may have a component of motion (*ticklish*, *swinging*, *scratchy*). And they may relate to gravity, such as declining (*low*), or being light (*downy*). Many of the words in this cluster also have a subjective feel to them, such as the words *ticklish*, *scratchy*, and *craggy*. However, these words are difficult to interpret as a coherent set because of their multisensory nature.

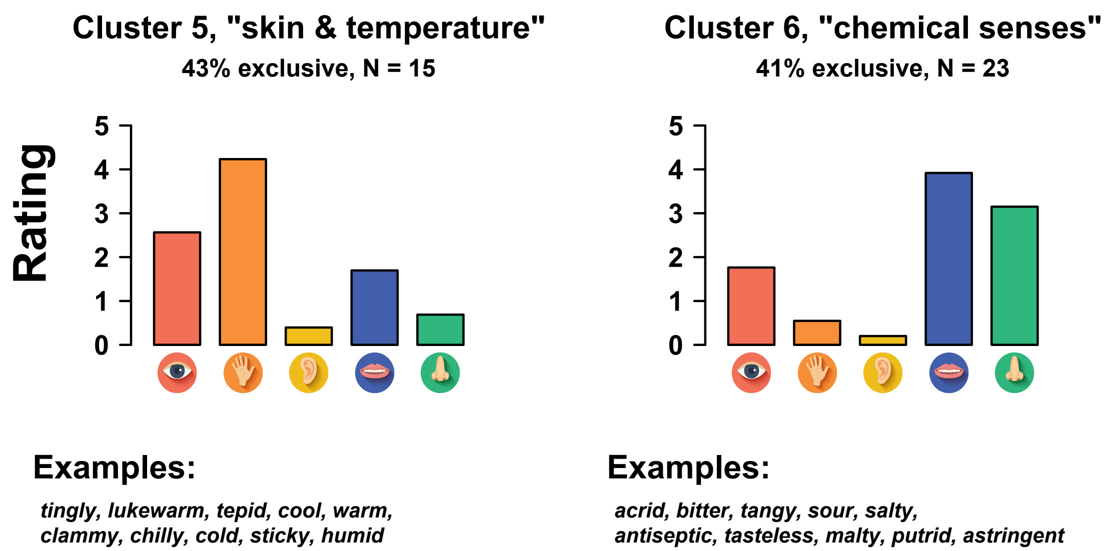


Figure 11. Clusters 5 and 6. Words are ranked by their certainty

Figure 11 shows the next two clusters. The "skin & temperature" cluster contains 15 words (average exclusivity 43%). The words in this cluster appear to describe properties that one can feel via one's skin, such as *tingly*, including temperature (*lukewarm*, *tepid*, *cool*, *warm*, *chilly*, *cold*). They also relate to liquidity and viscosity, such as *clammy*, *humid*, and *sticky*. The fact that words in this cluster have high touch ratings appears intuitive—we associate touch with the skin, and the skin is also the primary medium through which we experience temperature. It is remarkable that temperature-related words stick out from the other touch words. In some works on synesthetic metaphors, temperature is treated as separate from touch (Day, 1996; Ronga et al., 2012; Shinohara & Nakayama, 2011; Ullmann, 1945, 1957; Whitney, 1952); in others, touch and temperature are lumped together (e.g., Shen, 1997; Strik Lievers, 2015).

The way words in the "skin & temperature" cluster are differentiated from the other touch-related clusters is by having comparatively lower sight ratings and comparatively higher taste ratings. Although the visual ratings are still high, the fact that they are overall lower than for other touch-related clusters appears to correspond to the fact that temperature is something that is not always visually apparent. After all, we need warning lights on stoves precisely because it cannot easily be seen whether a surface is heated or not. But what explains the relatively high taste ratings of this cluster? We commonly experience the properties of temperature, liquidity, and viscosity in the context of food. This association with food may be what drives gustatory ratings up for this cluster. Perhaps participants in the rating study of Lynott and Connell (2009) used a heuristic whereby they judged all words related to food as gustatory.

The "chemical senses" cluster contains only 23 words (average exclusivity 41%). The words in this cluster describe properties that can be perceived through taste and smell. Properties such as *malty* and *antiseptic* may be experienced through ingestion or sniffing, and this may help explain why this cluster has both high gustatory ratings and olfactory strength ratings. Several of the words in this cluster have negative connotation (*acrid*, *bitter*, *antiseptic*, *putrid*, *astringent*) (see Chapter 16).

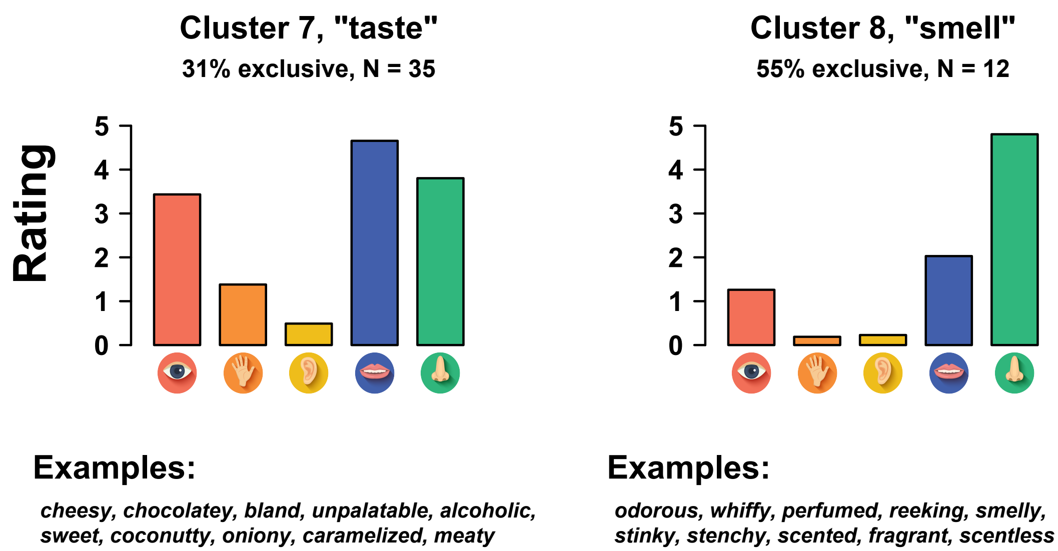


Figure 12. Clusters 7 and 8. Words are ranked by their certainty

Figure 12 shows the next two clusters. The "taste" cluster contains 35 words strongly relating to taste. The "smell" cluster contains 12 words strongly relating to smell. The taste cluster is more multisensory (exclusivity 31%) than the smell cluster (55%). This may, in part, be because the taste cluster is associated with food terminology, such as *cheesy*, *chocolatey*, *alcoholic*, *coconutty*, *oniony*, *caramelized*, and *meaty*. Whether something is *caramelized* or *chocolatey* can not only be tasted and smelled, it can often also be seen. This high multisensoriality of taste words also corresponds to the exclusivity analyses seen in Chapter 12 and the fact that taste words are frequently used in crossmodal expressions (Chapters 6–9).

Just as would be expected based on the Principal Components analysis earlier in this chapter, taste and smell words cluster together. However, there are also signs of asymmetry between the two senses: The chemical senses cluster has overall higher taste than smell ratings, and the smell cluster is much smaller than the taste cluster. This suggest that taste is more lexically differentiated than smell (Chapter 12). It should also be noted that whereas at least some of the taste words single out specific taste properties—such as *sweet*, *cheesy*, and *alcoholic*—the words in the smell cluster are much vaguer. As stated by Majid and Burenhult (2014, p. 266), terms such as "*stinky* or *fragrant* appear to denote the evaluative experience of the participant rather than the quality of the smell." Using a word such as *sweet*, even though it is also evaluative, gives a clear impression of a particular taste. The odor word *perfumed*, however, is only evaluative and has little precise perceptual content; this word could be used to describe many very different kinds of odors. These facts may be connected to the observation that "gustatory terms, such as sour, sweet, or pungent, usually double for olfactory terms" (Classen, 1993, p. 52; see also Classen et al., 1994, p. 109), as was already discussed in Chapter 8.

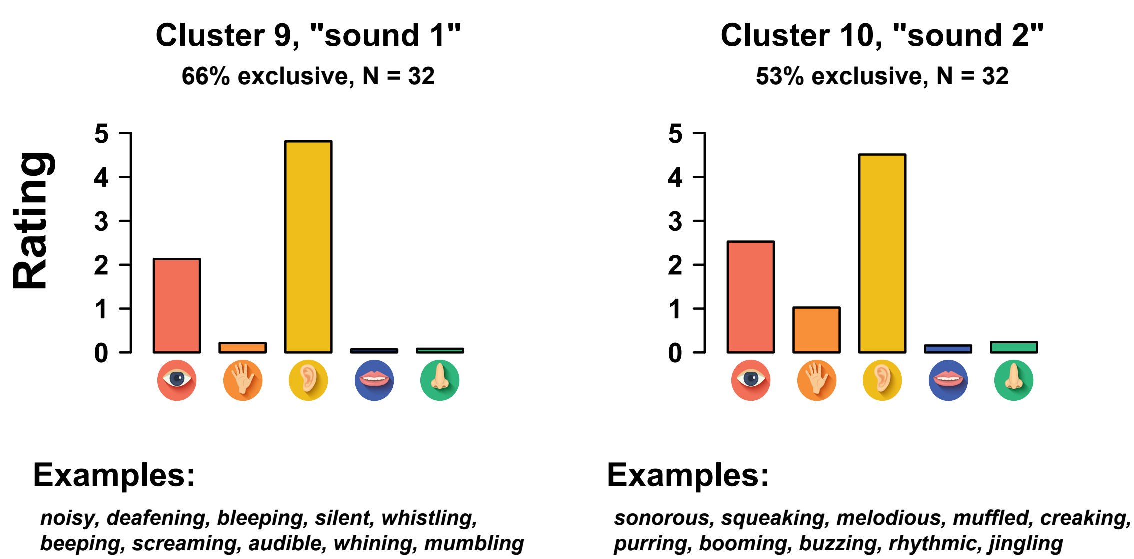


Figure 13. Clusters 9 and 10. Words are ranked by their certainty

Figure 13 shows the next two clusters, both of which relate to sound and contain 32 words each. Both sound clusters are relatively exclusive (66% and 53%). The difference between these two clusters is difficult to interpret. Perhaps the "sound 1" cluster is more related to sounds that machines (*beeping*, *bleeping*) or humans (*screaming*, *whistling*, *whining*, *mumbling*) would produce. Words within the "sound 2" cluster are sounds that animals can produce (*creaking*, *purring*, *squeaking*). Many of the properties within this cluster also appear to relate to music (*melodious*, *rhythmic*, *jingling*, *sonorous*). The properties in the "sound 1" cluster appear to be more directly measurable, referring perhaps more strongly to loudness and pitch. In contrast, the properties in the "sound 2" cluster appear to be more related to impressions or the effects of sounds on the perceiver. That said, the differences between these two sound clusters are particularly subtle when compared to some of the other clusters, which more obviously differ from each other.

Figure 14. Clusters 11 and 12. Words are ranked by their certainty

Figure 14 shows the final two clusters. The "impression-related" cluster contains 30 words and is relatively multisensory (43%). Words in this cluster appear to relate to impressions that can be gleaned primarily through vision, but also through the other modalities, such as *stormy*, *crackling*, *radiant*, *misty*, and *amber*. Some of the words also appear to characterize a certain level of intensity (*stormy*) or lack thereof (*mellow*).

The "multisensory" cluster has an average exclusivity of 28%, the lowest of all clusters. With 72 words it is also the biggest cluster. At least, words ranked highly in terms of certainty of cluster association (see Figure 14), appear to have a strong evaluative character—such as *beautiful* and *gorgeous,* which are very positive, or *grotesque, strange* and *dirty,* which are relatively more negative. Words in this cluster appear to be related to qualities that relate to our physical condition (such as being *dirty*, *clean*, *sweaty*) or our looks (*beautiful*, *gorgeous*, *grotesque*). Given its size (72 out of 423 words is 17% of the entire adjective dataset), this cluster is, however, a relatively mixed bag of words.

To conclude this section on cluster analysis, it is useful to relate the obtained clusters to the idea of ineffability discussed in Chapter 4. How many words are there within each cluster? This addresses Levinson and Majid's (2014) claim that ineffability is not necessarily about entire senses, but about perceptual qualities. We can interpret each cluster to be about a particular domain of experience. The number of words within each cluster can then be understood as an indicator of how differentiated the vocabulary for the expression of this domain is. Ordered in terms of word types, the "multisensory" cluster was the most differentiated (with 72 words), followed by "gross surface properties" (52), "pure sight" (41), "motion, touch, & gravity" (40), "shape & extent" (39), "taste" (35), "sound #1" (32), "sound #2" (32), "impression-related" (30), "chemical senses" (23), "skin & temperature" (15), and "smell" (12). These type counts reinforce the low degree of lexical differentiation for smells in English, especially in comparison to the "taste" cluster. These counts furthermore suggest that there is low lexical differentiation for "skin & temperature" words. The counts also suggest that a large number of sensory words characterizes the gross dimensions of the world we live in, including shape, extent, and gross surface properties. Space is immensely important for humans, and thus it comes as no surprise that the vocabulary for space is fairly differentiated.

The word type counts for the different clusters also reflect some degree of visual dominance: The "pure sight" cluster is quite lexically differentiated (with 41 words). The relatively large size of this cluster is particularly noteworthy because this cluster is also quite unisensory (very high average exclusivity). It is comparatively easier for multisensory clusters to be large in size. Moreover, many of the clusters that relate to surface properties and extent are rated to be highest in visual strength. This also suggests that English speakers appear to associate the spatial characteristics of our perceptual world most strongly with sight and less so with touch.

13.4. Revisiting the five senses model

In the Lynott and Connell (2009) rating study, participants were asked to rate each word with respect to each of the five sensory modalities. Despite this constraint, participants could not help but make more fine-grained distinctions, as well as form larger groupings. This serves to show that the constraining nature of the rating task does not prevent us from investigating more detailed structures, given the right statistical tools for uncovering said structures (compare Chapter 10). This is an important methodological point: Ratings can uncover conceptual structures that are not directly measured (see also Troche, Crutch, & Reilly, 2017). This is not only important for sensory linguistics, but also for the study of lexical semantics more generally. Lexical semanticists often group words (such as words for particular types of events or particular types of motion) based on either intuition or intuition-based linguistic tests (e.g., tests for aspectual classes). The methods presented in this section show that it is a worthwhile endeavor to also investigate which lexical groupings emerge from a bottom-up statistical approach when it is applied to rating data. Future research needs to extend this methodology to other semantic domains.

What do these results say about the five senses folk model? When zooming out to take a big picture perspective of the Lynott and Connell (2009) dataset, there is clear evidence for less than five groups, with taste/smell and touch/sight patterning together, both of which are separate from sound (Chapter 13.2). However, when zooming in, there is clear evidence for more than five groups, including groups that relate to temperature-related properties and groups that relate to the spatial dimensions of the environment. Just as the larger groupings, most of these smaller groupings crosscut the five senses, which is evidenced by the fact that there were relatively few unisensory clusters (such as the "pure sight" cluster and the two clusters related to sound).

Crucially, however, both the micro and the macro perspective suggest that the five senses model does not provide a full picture of English sensory adjectives. When looking at microstructures, it is necessary to draw more distinctions than just five. When looking at macrostructures, there are fewer groups than the five senses. Both of these perspectives are equally true. Clusters and correlations provide two complementary perspectives of the same dataset. One focuses on microstructures, the other one on macrostructures.

[1] Standardized PCA yields different results from unstandardized PCA and has been argued to improve the signal-to-noise ratio (Eklundh & Singh, 1993).

[2] In contrast to such heuristic techniques as *k*-means, Gaussian mixture modeling is a model-based technique where the data is assumed to come from a mixture of Gaussian probability densities. In this case, using mixture models is preferred over *k*-means clustering because it allows fuzzy associations to clusters. Mixture models are also less constraining in the geometrical shapes that are allowed for clusters. One should keep in mind, however, that different cluster analysis techniques will yield different cluster solutions. The clusters presented in this chapter thus are an abstraction that is specific to a cluster analysis using Gaussian mixture models.

[3] In a first run, I applied mixture models to the two-dimensional space that resulted from applying PCA to the raw modality norms. In this case, the data best supported an 8-cluster solution. If mixture models were run on the uncompressed raw data, the data supported a 12-cluster solution. The resulting clusters are qualitatively similar to the 8-cluster solution, but the 12-cluster solution turned out to be more interpretable. The fact that the additional clusters were readily interpretable suggests that they represent meaningful additional subgroupings. Thus, I chose to report the 12-cluster solution.

Chapter 14. Semantic preferences of sensory words

14.1. Introduction

An extensive number of studies show that sensory experience is fundamentally multisensory (O'Callaghan, 2015; Spence & Bayne, 2015) and there is a vast array of crossmodal interactions between different sensory modalities (Deroy & Spence, 2013; Spence, 2011, 2012, 2013). But when it comes to language about the senses, do all the senses work together equally well?

Chapter 12 reported noteworthy relations between the senses when looking at just the perceptual strength ratings. In this chapter, I will look at how these sensory relations pan out in naturally occurring text, following Firth's (1957, p. 179) famous credo that "you shall know a word by the company it keeps." When it comes to adjectives, the most relevant "company" are the head nouns they modify (e.g., Givón, 2001, p. 53; Miller & Johnson-Laird, 1976, p. 355). Adjectives are intrinsically underspecified, requiring a head noun to gain full meaning (Paradis, 2000). A sensory linguistic analysis may gain a deeper understanding of the sensory vocabulary of English by looking at how the senses are related with each other in adjective–noun pairs.

All the findings discussed in this chapter provide quantitative evidence for a core concept in corpus linguistics: the idea that words may have what is called "semantic preference" (e.g., Partington, 2004; Stubbs, 2001). Hunston and Francis (2000, p. 137) say that "a word may be said to have a particular semantic preference if it can be shown to co-occur typically with other words that belong to a particular semantic set." Semantic preference is related to the concept of collocation, which is "the statistical tendency of words to co-occur" (Hunston, 2002, p. 12). Semantic preference, then, is collocation at the level of concepts. We can use the term "modality affinity" to describe the semantic preference of sensory words in terms of which other senses they tend to co-occur with.

There are different ways of analyzing modality affinity. The first way pertains to the overall fit between the adjective and the noun; that is, do words with a particular sensory modality profile occur together with words of a related sensory modality profile? The second way of addressing the notion of modality affinity is by looking at correlations between particular sensory modalities. For instance, given the results presented in the last chapter, taste and smell words are expected to be used together in text. The following gives an overview of this chapter:

(a) cosine similarity analyses (Chapter 14.2)
(b) correlation analyses (Chapter 14.3)

Whereas the first set of analyses is targeted at gross modality affinity, disregarding specific sensory modalities, the second set of analyses is targeted at uncovering the inter-relations between the senses, looking at specific pairs of sensory modalities. Together, both analyses show that there is structure to how sensory words are used in natural language. Even though there is intense multisensoriality, it is not the case that "anything goes."

14.2. Cosine similarities

This section asks the question: Is there modality affinity for adjective–noun pairs? Do adjectives stick to nouns with highly related modality profiles? This puts the first and most general constraint on how sensory adjectives are used in context. To assess this, the Corpus of Contemporary American English (COCA, Davies, 2008) will be analyzed in conjunction with the Lynott and Connell (2009) adjective and the Lynott and Connell (2013) noun datasets (see Chapter 11). This large register-balanced corpus includes over 450 million words from magazine, news, academic writing, fiction, and spoken language.

All adjective–noun pairs containing one of the 423 adjectives from Lynott and Connell's study were extracted. This yielded nearly 150,000 adjective-pair types (N = 149,387), with over a million adjective–noun pair tokens (N = 1,023,851). To investigate modality affinity, the adjective norms will be related to the noun norms from Lynott and Connell (2013). All adjective–noun pairs that did not feature a noun from Lynott and Connell's (2013) list were excluded, which resulted in a dataset that contained about 14,000 unique adjective-noun pair types (N = 13,685; 183,533 adjective-noun pair tokens).

Cosine similarity (see Chapter 11) can be used to quantify the degree to which an adjective's sensory modality profile is similar to a noun's sensory modality profile. The formula for cosine similarity is repeated here.

(E2) $$\text{pair similarity} = \cos(\theta) = \frac{adj*noun}{\|adj\|*\|noun\|}$$

The vectors represented by "adj" and "noun" in the above formula are the modality vectors of the two words being compared. A word can be conceptualized as a vector that is located within a five-dimensional modality space, defined by its ratings. For example, a highly sight-related adjective such as *hazy* points strongly into the direction of sight. In the space spanned by the five senses, adjectives and nouns with dissimilar modality profiles point in different directions. Adjectives and nouns with similar modality profiles point in similar directions. The similarity of two words can thus be quantified by the angle between the two vectors (using the cosine). The cosine similarity ranges from 0 (adjective and noun are maximally different, vectors are exactly perpendicular to each other) to 1 (adjective and noun are maximally similar, vectors are parallel to each other). The cosine thus quantifies the fit between an adjective's modality profile and a noun's modality profile.

Table 8 shows two adjective–noun pairs, *abrasive contact* and *sweet music*, with their corresponding perceptual strength values and the resulting cosines. It is instructive to compare the perceptual strength rating of each adjective with each noun. This shows that *abrasive* and *contact* share high sight and touch ratings, and they are also both low on sound, taste, and smell ratings. Correspondingly, the cosine similarity of this adjective–noun pair is very high (0.91). In contrast, the adjective *sweet* has a very different modality profile from the noun *music*. Whereas *sweet* has high ratings for taste and smell, *music* has high ratings for sound. As a result, the cosine similarity value for this adjective–noun pair is much lower (0.31). The difference in cosines between the two pairs indicates that the modality fit between *abrasive* and *contact* is higher than between *sweet* and *music*.

Table 8

Modality profiles of the adjective-noun pairs 'abrasive contact' and 'sweet music' with the corresponding cosine similarity

	Sight	**Touch**	**Sound**	**Taste**	**Smell**	**Similarity**
abrasive	2.89	3.68	1.68	0.58	0.58	
contact	3.41	3.53	2.53	1.06	1.12	**0.91**
sweet	2.19	0.57	1.19	4.86	3.90	
music	2.24	1.24	4.94	0	0.06	**0.31**

The cosine similarity between the adjective and the noun was computed for all 14,000 adjective–noun pairs that were attested in COCA. The average cosine similarity was 0.81. This figure is relatively far away from the lowest possible cosine similarity of zero. Thus, the figure 0.81 indicates that adjectives combine with nouns that have similar modality profiles. This is quantitative evidence for modality affinity.

However, we have to rule out that the figure of 0.81 is not mathematically inevitable. It should be kept in mind that Chapter 12 found many words to be high in sight ratings and low in taste, smell, and sound ratings. Given this, it is entirely possible that simply combining adjectives and nouns at random would lead to a high cosine similarity. In particular, the following question has to be asked: Is a cosine value of 0.81 actually higher than what is expected by chance? In essence, we need a statistical baseline for the figure of 0.81. Only if this empirically established cosine similarity is higher than a chance baseline is the concept of modality affinity supported by this data.

To address these concerns, I compared the difference in modality fit between those adjective–noun pairs that are attested in COCA to those adjective–noun pairs that are hypothetically possible, but that are actually unattested. The set of 14,000 adjective–noun pairs analyzed above is only a small subset (8%) of the set of possible adjective–noun pairings that could be generated with the two sets of norms. In particular, with 423 adjectives and 400 nouns from the two rating studies, there are at least 423 * 400 = 169,200 possible combinations. For instance, the pairs *laughing liquid*, *garlicky promotion*, *scratchy fortune*, *solid spirit*, *forked provision*, *thumping welcome*, *greasy sergeant*, *shrieking money*, and *moldy grace* are part of the 92% of possible pairs (N = 155,515) that are not attested in COCA. These unattested pairs serve as a hypothetical baseline for comparison with the actually attested pairs. Figure 15 shows the distributions of the attested (a) and unattested (b) pairs.

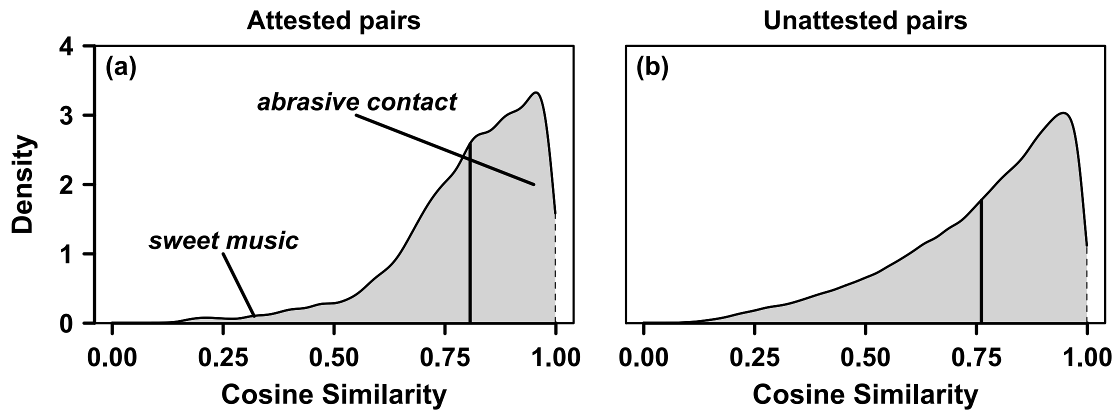


Figure 15. Kernel density estimates (over word types) as a function of cosine similarity of (a) adjective–noun pairs that are attested in COCA (N = 13,685) and (b) unattested adjective–noun pairs (N = 155,515); solid lines indicate means; density curves are restricted to observed range.

As can be seen in Figure 15, both the attested and unattested distributions have negative skew, with the tail tapering off toward lower values. That is, both distributions have many data points with relatively high cosines. This shows that indeed, given the distributional structure of the modality ratings, a high cosine similarity is to be expected (this may be, for example, because most adjectives and nouns are relatively high in visual content). However, crucially, the cosines of the attested and unattested pairs are reliably different from each other (Wilcoxon rank sum test with continuity correction, W = 931,630,000, p < 0.0001), with the average cosine similarity of the unattested adjective–noun pairs being 0.76, lower than the average cosine similarity of the attested pairs (0.81). Thus, the actually observed adjective–noun pairs have a higher modality fit than what would be expected given all possible combinations of adjective and noun pairs from the two datasets. This provides quantitative evidence for the idea that sensory words tend to combine with words that have similar perceptual characteristics.

14.3. Correlations within adjective–noun pairs

14.3.1. Predictions

So far, this chapter has looked at the overall modality fit between adjectives and nouns, ignoring differences between the senses. I now turn to looking at pairwise correlations. Which specific senses correlate with which other specific senses in natural language data?

First, Chapter 14.3.1 will generate predictions following from the Embodied Lexicon Hypothesis (Chapter 5). Then, Chapter 14.3.2 will test those predictions. In particular, I argue that two predictions can be made for crossmodal relations in adjective-noun pairs: First, taste and smell should associate with each other. Second, sight and touch should associate with each other. Previous chapters have touched on these specific crossmodal connections. This chapter will discuss taste/smell and touch/sight integration in more detail.

There is a wealth of evidence for taste/smell integration. Eating necessarily involves smelling (Mojet, Köster, & Prinz, 2005), in part because humans not only smell through the nose, the so-called "orthonasal pathway," but also through the so-called "retronasal pathway," a passage to the olfactory bulb at the back of the oral cavity. Smell perceived through both channels interacts with taste in determining flavor. Behavioral experiments found, for example, that caramel odor can suppress the sour taste of citric acid (Stevenson, Prescott, & Boakes, 1999), or that flavor and smell intensities and detection thresholds change depending on how taste and smell are combined (Dalton, Doolittle, Nagata, & Breslin, 2000; Delwiche & Heffelfinger, 2005; Pfeiffer, Hollowood, Hort, & Taylor, 2005). Taste and smell are also neurally integrated, sharing overlapping brain networks (de Araujo, Rolls, Kringelbach, McGlone, & Phillips, 2003; Delwiche & Heffelfinger, 2005; Rolls, 2008). In fact, taste and smell are so highly integrated and mutually interdependent that it is legitimate to ask whether they are actually distinct senses (see discussion in Spence et al., 2015). If taste and smell *words* follow these perceptual patterns, they should combine together in text.

What is the evidence for touch/sight integration? Touching generally also involves seeing (Walsh, 2000). Actions such as manually reaching for an object involve a concerted interplay between vision and touch. Determining shape via touch appears to involve visual mental imagery (Klatzky, Lederman, & Reed, 1987). Throughout human development, touch calibrates the size perception of vision, and vision calibrates the orientation perception of touch (Gori, Del Viva, Sandini, & Burr, 2008; Gori, Sandini, Martinoli, & Burr, 2010).

There is abundant evidence for neural integration of sight and touch as well: The occipital cortex, specifically the parietal-occipital fissure, shows increased blood flow when participants make visual judgments as well as tactile judgments of the orientation of grating patterns on surfaces (Alivisatos, Jacobson, Hendler, Malach, & Zohary, 2002; Sathian & Zangaladze, 2002; Sathian, Zangaladze, Hoffman, & Grafton, 1997; Sergent, Ohta, & MacDonald, 1992). Furthermore, Zangaladze and colleagues used transcranial magnetic stimulation over the occipital portion of the scalp and showed that this interfered with the tactile discrimination of grating orientation (Zangaladze, Epstein, Grafton, & Sathian, 1999). A previous study has shown that stimulation via a magnetic coil over the same region also interferes with the visual recognition of letters (Amassian et al., 1989). The intraparietal sulcus shows increased blood flow when performing mental rotation in both the visual domain and the tactile domain (Cohen et al., 1996; Prather, Votaw, & Sathian, 2004). More generally, large regions of the visual cortex also respond to somatosensory stimuli (Casagrande, 1994; Haenny, Maunsell, & Schiller, 1998; Hagen et al., 2002). If touch and sight *words* follow these neurophysiological and behavioral patterns, they should combine together in text.

Ronga and colleagues (2012) state that the usage of sensory words in context "follows the same tendencies governing the perceptual integration of different sensory modalities" (p. 155). According to Marks (1978), "interrelations among the senses that appear in perception will also find their way into speech and writing" (p. 3). These ideas will be explored here by showing there is a semantic preference for taste words to associate with smell words (and vice versa), as well as for sight words to associate with touch words (and vice versa).

Before I continue, the following question must be asked: Why not make similar predictions for other crossmodal interactions? Given the myriad of possible interactions between the senses (as reviewed in Spence, 2011), why stop at taste/smell and touch/sight integration? For instance, a pattern of crossmodal integration between sight and sound is also particularly dominant. There is abundant evidence for audiovisual integration (see Spence, 2007), just as is the case for taste/smell and touch/sight. For example, the McGurk-MacDonald effect shows that when participants see a video of somebody speaking, their auditory speech perception is affected by the precise execution of the lip and jaw movements (McGurk & MacDonald, 1976). The ventriloquist effect shows that vision can "pull" audition toward a particular spatial percept (Alais & Burr, 2004; Pick, Warren, & Hay, 1969; Welch & Warren, 1980). On the other hand, temporal ventriloquism shows that audition can pull vision toward a particular temporal percept (Morein-Zamir et al., 2003). The sound-induced flash illusion furthermore shows that hearing two beeps makes participants see a single light flash as two light flashes (Shams, Kamitani, & Shimojo, 2002). Audition and vision are also neurally integrated (e.g., Baumann & Greenlee, 2007). Given these manifold interactions between sight and sound, shouldn't the corresponding sensory words be associated with each other as well?

There are at least three reasons that speak against the prediction that audiovisual integration should lead to a predominance of audiovisual language. It is insightful to contrast audiovisual integration with the case of taste/smell and touch/sight interactions. First, objects that look rough or smooth also generally feel rough or smooth. The same goes for objects that smell sweet, which often (but not always) taste sweet as well. That is, the integration of touch/sight and taste/smell is bound to objects. The same cannot be said for sound, which is event-based, not object-based (see Strik Lievers & Winter, 2018). When someone uses the adjective *barking* in a description, such as *I heard a barking sound*, this barking "property" is a less stable characteristic of the object. Philosophers since Aristotle have emphasized the inherent dynamicity of sound in our phenomenology (O'Callaghan, 2009; O'Callaghan & Nudds, 2009). Sound language may be different from the other modalities by virtue of the inherent dynamicity of sound. Moreover, precisely because sound-related adjectives such as *squealing*, *barking*, *beeping*, *howling*, and *whistling* describe events rather than properties, these adjectives are likely going to be used together with different nouns, compared to other sensory words.

A second reason for not predicting a sight/sound association is that words such as *shrill*, *mumbling*, and *thumping* have an onomatopoetic feel to them, a topic that I will revisit in Chapter 15 (see also Perlman et al., 2018; Winter et al., 2017). As was already mentioned in Chapter 9.2.5, the iconicity of these words may tie them more strongly to the auditory modality, which is another factor that makes sound words more exclusive in usage (see also discussion in Lynott & Connell, 2013; Lupyan & Winter, 2018).

Together, the two reasons stated (the lack of stable object binding and iconicity in word formation) explain why I predict linguistic touch/sight and taste/smell interactions, but not sight/sound interactions—despite evidence for sight and sound interacting in perception.

Thus, I make three predictions for the following correlation analysis: First, sight-related adjectives should modify touch-related nouns and vice versa. Second, taste-related adjectives should modify smell-related nouns and vice versa. Third, sound words should be exclusive; they should not pattern together with words from the other senses.

14.3.2. Correlation analysis

This section re-uses the 14,000 adjective–noun pairs from Chapter 14.2—that is, the adjective–noun pairs from COCA for which both adjective and noun modality data exists. For each adjective and each sensory modality, the mean perceptual strength of the noun contexts was computed. For example, the word *abrasive* modifies the nouns *contact*, *dust*, and *paper*, among many other nouns). These three nouns have visual strength ratings of 3.4, 4.2, and 4.4, respectively. For the example of these three values only, this yields an average noun visual strength of 4.0 for the adjective *abrasive*. Such averages were computed in a frequency-weighted fashion (i.e., adjective–noun pairs with higher token frequencies contributed more towards the mean).

Figure 16 visualizes the correlations between adjectives and nouns. The direction of the arrows is to be interpreted as follows: An arrow that points from sight to touch, for instance, describes the correlation between the visual strength of the adjective and the tactile strength of the noun (in this case, r = 0.37). Conversely, an arrow pointing from touch to vision describes the correlation between the tactile strength of the adjective and the visual strength of the noun (in this case, $r = 0.33$). The figure only shows correlations which were indicated to be reliably different from zero after Bonferroni correction for performing 25 tests ($p < 0.002$). For the sake of simplicity, the visual representation omits one statistically reliable negative correlation (between an adjective's visual strength rating and a noun's taste rating).[1]

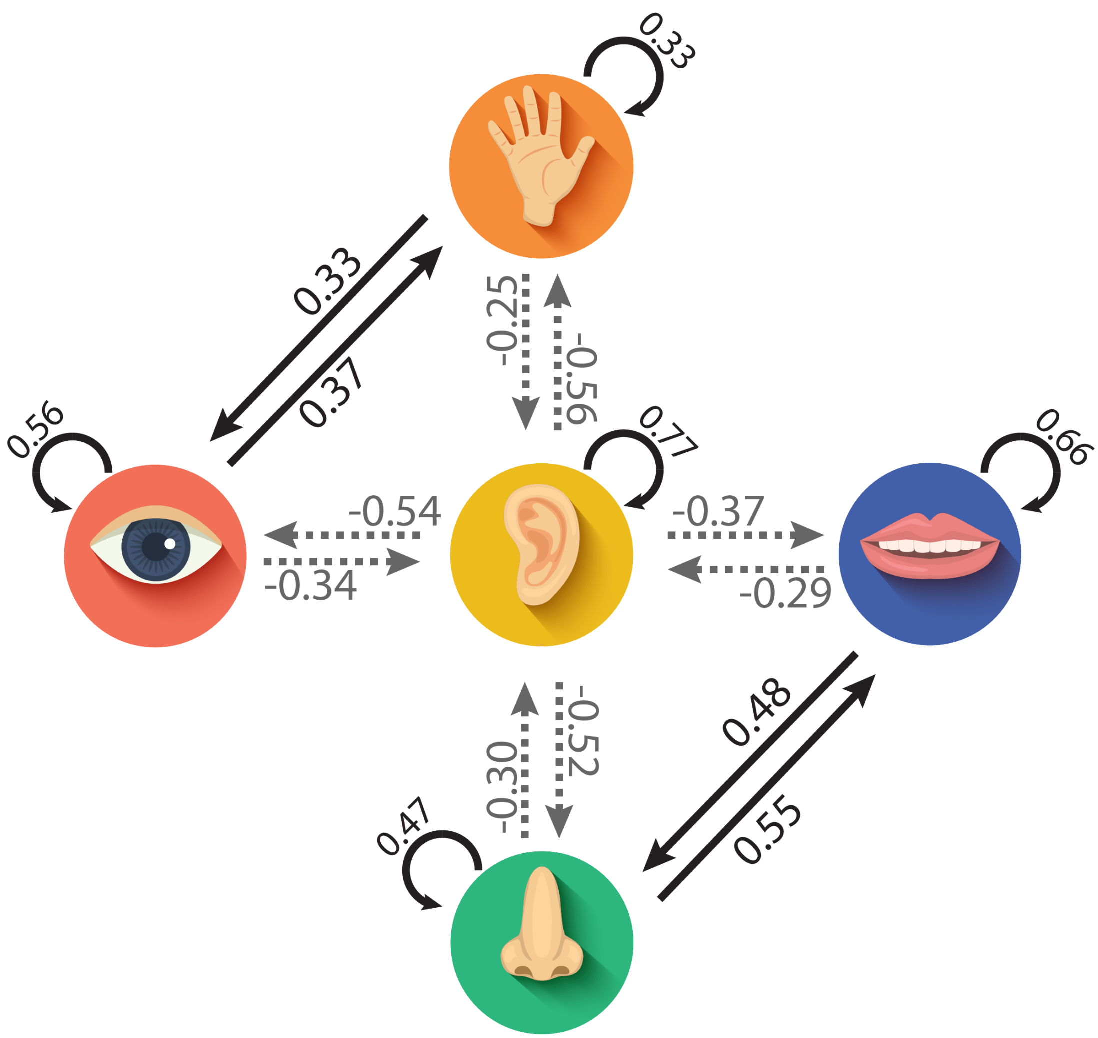


Figure 16. The correlational structure of multisensoriality. Data from 13,685 adjective-noun pairs; solid arrows indicate statistically reliable correlations (corrected for performing 25 comparisons), dotted arrows indicate statistically

reliable anti-correlations; the arrow heads point "from the adjective to the noun," i.e., the vision-to-touch arrow indicates that the visual strength of an adjective is, on average, correlated with the tactile strength of the noun with $r = 0.37$. A weak negative correlation between an adjective's visual strength and the gustatory strength of the noun ($r = -0.22$) is not shown in this figure.

First, let us focus on the within-modality associations shown in Figure 16. For each modality, there were statistically reliable correlations with itself. This means that adjectives like to pair with nouns that have high perceptual strength ratings for the same modalities. This finding corroborates the cosine analyses reported above (Chapter 14.2). Moreover, it should be noted that all of the correlations are far from 1.0, the highest possible correlation. While part of this could be due to measurement error (a word's modality rating cannot be measured perfectly, which is going to drive correlations down), this can also be interpreted as evidence for multisensoriality in adjective–noun pairs: Adjectives do not *only* go together with nouns from the same modality, they are also used crossmodally. Moreover, the degree to which adjectives are used crossmodally differs between the dominant modality considered: Touch has the lowest intra-modal correlation ($r = 0.33$), suggesting that touch words are most likely used in contexts that are outside of their own modality. This is followed by smell ($r = 0.46$), sight ($r = 0.56$), taste ($r = 0.66$), and sound ($r = 0.77$). The fact that sound has the highest intra-modal correlation fits the predicted pattern of auditory exclusivity: Sound words are most likely to be used in the context of other sound words.

Next, let us look at the correlation between sight and touch. The fact that there are arrows pointing both ways means the following: First, sight adjectives modify nouns that can also be felt, such as is the case with *shiny belt*, *shiny body*, and *shiny glass*, all of which are attested adjective–noun pairs. Second, touch adjectives modify nouns that can also be seen, such as *rough blanket*, *rough cotton*, and *rough landscape*. A similar bidirectional relationship characterizes taste and smell words.[2] For example, the highly olfactory word *smoky* (which is also quite gustatory) occurs in such expressions as *smoky taste*, *smoky food*, and *smoky sauce.* Similarly, the highly gustatory word *sweet* occurs in such highly olfactory pairs as *sweet whiff*, *sweet rose*, *sweet balsam*, and *sweet cologne*.

Finally, Figure 16 clearly shows that sound is anti-correlated with everything else. This means that sound adjectives are not frequently used to modify nouns from the other senses. The auditory adjectives *squealing*, *booming*, and *muffled*, for instance, tend to modify such auditory nouns as *sound* and *music*, and not nouns such as *sauce* (taste), *cotton* (touch), and *picture* (sight). Moreover, adjectives from the other sensory modalities do not frequently modify sound nouns. The words *music* and *sound*, for instance, are predominantly described with auditory adjectives and much less so with non-auditory adjectives.[3]

14.4. The structure of multisensoriality

In this chapter, I have presented several results that point to the idea that words with similar modality profiles stick together. This result was established for adjective–noun pairs only, but similar principles are expected to hold for other types of sensory expressions as well. Chapter 14.2 showed that compared to unattested adjective–noun pair types, attested pairs were characterized by a strong modality fit. Moreover, those adjective–noun pair types with high modality fit had higher token frequencies. Chapter 14.3 furthermore showed that taste/smell and touch/sight words are associated with each other in adjective–noun pairs. Together, these findings provide quantitative evidence for the corpus-linguistic notion of semantic preference extending to the domain of the senses. Words preferentially combine either with words from their own sensory modality, or with words from highly related modalities. This phenomenon was referred to here as modality affinity, a subtype of semantic preference.

The only modality that stood out from the rest was sound, which was found to be anti-correlated with all other senses (Chapter 14.3). Words such as *warbling*, *hoarse*, and *growling* are perfectly fit for describing sound sensations, but they are much less apt for describing other sensory phenomena. Similarly, highly auditory nouns such as *laughter*, *voice*, and *harmony* are not frequently described using non-auditory words such as *yellow*, *oniony*, or *odorous*.

Together, these findings suggest that the multisensoriality of sensory words is structured. It is not the case that "anything goes" with respect to the distributional patterns of sensory words. There are clear indications of affinity between some senses—such as taste/smell and sight/touch—but also clear indications of aversion between others, particularly between sound and the rest. This structure was in part predicted by the Embodied Lexicon Hypothesis (Chapter 5) based on the fact that taste/smell and touch/sight are also associated with each other in perception. Knowing about how the senses play together in perception allows predicting, at least to some extent, how the corresponding sensory words behave in language. That is, linguistic association is predictable from perceptual association.

The results of this chapter also have important implications for corpus linguistics. First, the analyses used here showcase a new methodological approach to quantifying the notion of semantic preference. In particular, it is possible to combine data collected in a decontextualized rating task with corpus data to quantify the degree to which words of similar semantic profiles stick together, in this case words of similar sensory modality profiles. Second, the results from this chapter need to be compared to the results of the preceding chapter. In particular, Chapter 13 uncovered crossmodal relations that looked much like what was found in the corpus analysis of this chapter: Chapter 13 showed that taste and smell ratings were correlated with each other, and so were touch and sight ratings. Very similar relations were found when looking at adjective–noun pairs in corpora. This similarity in findings, in fact, shows that native speaker intuitions obtained in a decontextualized task meaningfully correspond to how sensory words are used in context.

The close correspondence between decontextualized ratings and contextualized usage is an important result because corpus linguists generally argue against an over-reliance of native speaker intuitions. For example, Hunston (2002) states that "...the main argument in favour of using a corpus is that it is a more reliable guide to language use than native speaker intuition is" (p. 20). This statement may well be true, but this chapter has presented evidence demonstrating what people think about language in a decontextualized task is predictive of contextual use. This is in line with Miller and Charles (1991), who showed that similarity ratings on isolated word pairs were predictive of corpus-based contextual similarity. The close correspondence of the results in this chapter with the results of the preceding chapter shows that sensory words can be meaningfully studied in isolation as well as in context, and the two analyses can inform each other.

To conclude, this chapter revealed clear structure to the crossmodal ways with which sensory words are employed by speakers of English. To use William James's description, there is no "great blooming, buzzing confusion" of sensory words; but the analyses of this chapter also show that there are no crisp divisions between the senses either. Sensory words are neither fully constrained nor fully flexible with respect to the contexts in which they occur. Sensory words can be said to occupy a sweet spot of multisensoriality.

[1] Adjective visual strength ratings were anti-correlated with noun gustatory strength ratings ($t(394) = 4.4$, $p < 0.001$), with a negative Pearson's r of –0.22. This reveals that sight-related adjectives are not used frequently in highly taste-related contexts. This is perhaps surprising because visual descriptors and color terms such as *yellow* can clearly be used in food-related contexts, such as the following expressions that occurred in the corpus: *yellow food*, *yellow liquid*, and *yellow sauce*. However, sight-related words appear much more frequently in contexts that have nothing to do with taste, such as *yellow shirt*, *yellow hat*, and *yellow eye*. Clearly, English speakers use visual words in the context of food to describe how food looks, but the frequency of these food contexts does not outweigh the frequency of non-food contexts. This may be the reason why the visual strength of the adjective is anti-correlated with the gustatory strength of the noun. A similar view is, in fact, expressed by Ronga (2016):

> When we eat, we perceive the taste of our food and simultaneously its fragrance and texture, but we are not able to look at the food in our mouth. This may be the reason why synaesthetic pairings composed of visual modifiers are rare. (p. 57)

However, one should also not overstate this result because it is the weakest of all the correlations found.

[2] Taste and smell are often co-lexified on the same lexical items also in other languages. For example, some varieties of German (such as Swiss German) use the same word *schmecken* to describe the experience of tasting and the experience of smelling.

[3] It is noteworthy that all correlations 'from' sound to the other modalities (sound adjectives modifying other-modality nouns) are lower, rather than the other way around. This pattern conceptually relates to what is observed in synesthetic metaphor research (Chapters 6–9; Chapter 17), where sound concepts are particularly infrequent as descriptors of other sensory perceptions.

Chapter 15. Frequency, semantic complexity, and iconicity

15.1. Introduction

Up to this point, every chapter has exclusively dealt with the modality norms and the Corpus of Contemporary American English (Chapter 14). This means that we have almost exclusively looked at inter-relations between the five senses. To further understand the sensory vocabulary of English, the list of sensory adjectives needs to be related to other linguistic patterns. This will be done here, where I use word frequency data (Chapter 15.2), dictionary meaning counts (Chapter 15.3), and iconicity norms (Chapter 15.4) to gain a deeper understanding of the sensory vocabulary of English.

I will also continue the theme of exploring the extent to which the five senses folk model is supported by linguistic data (Chapter 2). Throughout this chapter, I will repeatedly perform three types of statistical analyses in parallel. For each of the different datasets (word frequency, emotional valence, etc.), I will assess whether they are best described in terms of any of the three following types of models:

(a) **Cluster model:** Regression of the dependent variable (frequency, valence, etc.) onto the factor Cluster, which comprises 12 levels (one for each cluster from Chapter 13)

(b) **Categorical model:** Regression of the dependent variable onto the factor Dominant Modality, which comprises five levels taken from the modality norms

(c) **Continuous model:** Regression of the dependent variable onto the five continuous perceptual strength measures taken from the modality norms

The comparison of these three models allows us to assess which way of carving up the sensory world best accounts for linguistic patterns. How far can we take the five senses folk model? How much is gained by taking finer-grained distinctions into account? Finally, are sensory words best treated in a categorical or a continuous fashion?

15.2. Word frequency

Chapter 12 found that vision was the most differentiated sense, having the most unique word types. I will now explore whether visual concepts are verbalized more frequently. Type and token frequency are logically independent from each other (Regier et al., 2016; Warriner & Kuperman, 2015).

Some evidence already suggests that English speakers verbalize visual concepts more frequently (San Roque et al., 2015; Viberg, 1993), but this has so far only been shown for perception verbs such as *to see*. Does visual dominance in word frequencies extend to sensory adjectives? In this section, I will repeat some of the analyses conducted in Winter et al. (2018). The analyses will be extended to allow comparison of the three different models discussed above.

I will analyze word token frequencies from the SUBTLEX corpus of American English (Brysbaert & New, 2009). This corpus has approximately 51 million words from over 8,000 different American English movies and TV shows.[1] SUBTLEX word frequencies have been argued to correspond particularly well to human behavior, such as reaction times in psycholinguistic experiments (see Brysbaert & New, 2009). Four hundred thirteen words from the Lynott and Connell (2009) dataset were also represented in SUBTLEX (98%).

Figure 17 depicts a boxplot of the logarithmically transformed frequencies (log10) for each of the clusters introduced in Chapter 13. Boxplots are colored in terms of the sense that had the highest average ratings. It should be kept in mind, however, that several of the clusters were highly multisensory.

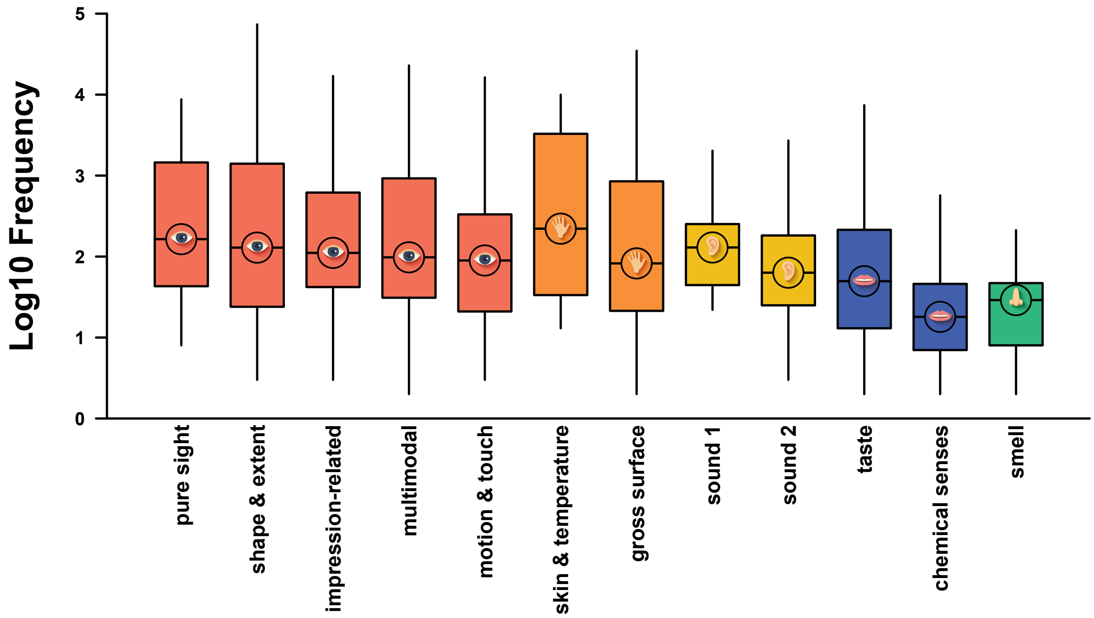


Figure 17. Log10 frequency from SUBTLEX by cluster**.** Boxplots are colored in terms of the dominant modality for each cluster and ordered by modality and frequency within each modality. Clusters are taken from Chapter 13. The boxes cover 50% of the data, with the middle line indicating the median. The whiskers cover the largest or smallest value within the interval of 1.5 times the interquartile range.

A look at Figure 17 suggests that sight, touch, and sound words are overall quite frequent in the SUBTLEX corpus. Words within the “skin & temperature” cluster are also relatively more frequent than words in the “gross surface” cluster. There appear to be no major frequency differences between the different visual clusters. On the other hand, taste and smell words occur less frequently in SUBTLEX. There also is a notable frequency asymmetry between the taste cluster, which is more frequent than the clusters for the chemical senses and the cluster for smell.

A simple one-way ANOVA with the 12-level Cluster predictor (cluster model) revealed a reliable effect on log frequencies ($F(11, 401) = 3.3$, $p = 0.0002$), which described about 6% in variance (adjusted $R^2 = 0.06$).[2]

How does the cluster model compare to a model that predicts frequencies based on the dominant modality classifications? A simple one-way ANOVA with the 5-level Modality predictor (categorical model) revealed a reliable effect ($F(4, 408) = 7.1$, $p < 0.0001$), which also described about 6% in variance ($R^2 = 0.06$). Sight adjectives were the most frequent (log mean = 2.2), followed by touch ($M = 2.1$), sound ($M = 1.9$), taste ($M = 1.6$), and smell ($M = 1.6$). In terms of raw frequencies, sight adjectives occurred on average 2,018 times in the corpus. Touch words occurred 1,036 times. Sound words occurred 347 times. Taste words occurred 267 times. And smell words occurred 248 times.

How do these two analyses fare against the continuous perceptual strength ratings from each modality? A multiple regression model was fitted on log frequencies with five continuous predictors, one for each modality.[3] Taken together as a set, the continuous perceptual strength ratings reliably predicted log frequencies ($F(5, 407) = 13.3$, $p < 0.0001$) and described 13% of the variance ($R^2 = 0.13$). A look at the regression coefficients revealed positive frequency slopes for all perceptual strength measures except for smell. Only the sight (estimate: +0.32, $SE = 0.05$, $p < 0.0001$) and sound (+0.14, $SE = 0.03$, $p < 0.0001$) slopes were reliably different from zero.

Finally, how do the cluster model, the categorical model, and the continuous model compare? The relative performance of each statistical model to the frequency data can be assessed using Bayesian Information Criterions (BICs). This measure balances model complexity and model fit. Relatively lower BIC values indicate better model performance.[4] The continuous model had the lowest BIC (1102), followed by the categorical five senses model (1130), followed by the clusters model (1164). Thus, BICs suggest that the continuous model performed best.

The cluster model and the categorical model both regressed frequencies onto a categorical data structure. The two models accounted for a similar amount of variance, but the categorical model did so in a much more parsimonious fashion, using five rather than twelve distinctions. The relatively poor performance of the cluster model is also apparent in Figure 17: The different clusters within each dominant sense did not differ much in terms of word frequency, with the notable exception of the "skin & temperature" cluster and the "taste" cluster.

15.3. Dictionary meaning counts

In this section, I will use dictionary meaning counts to test the idea that different classes of sensory words differ in terms of their semantic complexity. Sight and touch (and perhaps sound as well) should be higher in semantic complexity than taste and smell. This is to be expected for several reasons. First, frequency is correlated with the number of dictionary meanings (Baker, 1950; Köhler, 1986; Thorndike, 1948; Zipf, 1945, 1949), and speakers prefer to semantically extend frequent as opposed to infrequent forms (Harmon & Kapatsinski, 2017). Therefore, any differences in word frequency (Chapter 15.2) should be associated with differences in semantic complexity.

Moreover, Viberg (1983) and Evans and Wilkins (2000) have argued that sight, touch, and sound are relatively more prone to semantic extension than taste and smell (see also Sweetser, 1990).[5] For example, metaphors for intelligence frequently derive from the tactile modality, such as describing somebody as *acute*, *keen*, *sharp*, or as having a *penetrating* mind (Classen, 1993, p. 58; Howes, 2002, pp. 69–71). Others have stated that sight is particularly prone to acquiring metaphorical meanings denoting mental content (Caballero & Ibarretxe-Antuñano, 2014; Caplan, 1973; Matlock, 1989; Sweetser, 1990; though see Evans & Wilkins, 2000; see also San Roque, Kendrick, Norcliffe, & Majid, 2018), as in the English expression *I see* meaning 'I understand'. The same goes for many verbs of hearing, such as *I hear you* also meaning 'I understand' (see Sweetser, 1990, p. 41). We may also expect touch to be particularly high in semantic complexity because of the observation that it is a frequent source domain in crossmodal language use (Chapter 8), as in *sharp pitch*, *rough voice*, *abrasive tone*, and *smooth melody*.

I used dictionary meanings from WordNet (Fellbaum, 1998; Miller, 1995). Dictionary meaning data was available for 370 of the 423 adjectives (87%). Although dictionary meanings do not directly reflect the semantic structure of the mental lexicon (e.g., Croft & Cruse, 2004; Elman, 2004), they have been shown to correspond meaningfully to psycholinguistic behavior (Gernsbacher, 1984; Jastrzembski & Stanners, 1975; Johnson-Laird & Quinn, 1976; Jorgensen, 1990), and they have successfully been used as shorthand for semantic complexity in other works (e.g., Zipf, 1945, 1949; see also Baayen & del Prado Martín, 2005).

Figure 18 shows a boxplot of dictionary meaning counts per cluster. The shape of the boxplot clearly reveals positive skew, which is to be expected from a count variable (the data was additionally analyzed with negative binomial regression to show that the results are robust; see online supplementary materials). Logarithmically transformed sense counts (log10) were regressed on the Cluster factor, while simultaneously controlling for frequency. This analysis revealed a reliable effect of Cluster ($F(11, 354) = 5.4$, $p < 0.0001$; $R^2 = 0.07$). In addition, there was a reliable effect of Log Frequency ($F(1, 354) = 285.5$, $p <$

0.0001), thus replicating the known finding (e.g., Köhler, 1986; Zipf, 1945, 1949) that frequent words have many dictionary meanings (unique multiple $R^2 = 0.41$).

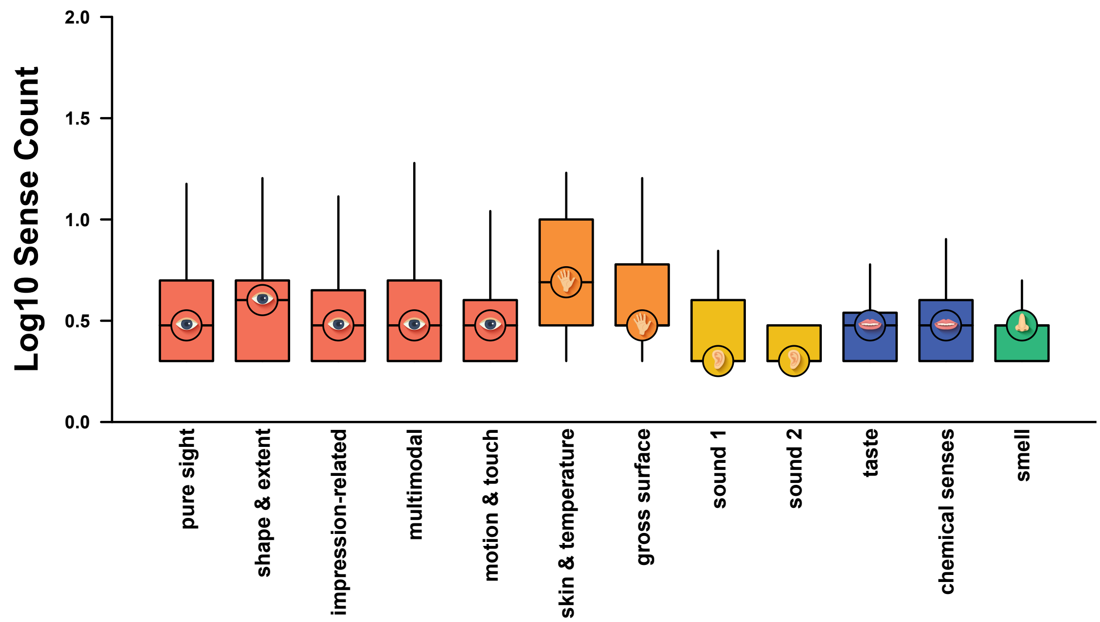


Figure 18. Log10 dictionary meaning counts by cluster.

Dictionary meaning counts also differed reliably by dominant modality while simultaneously controlling for Log Frequency ($F(4, 361) = 11.84$, $p < 0.0001$; $R^2 = 0.04$). On average, touch words had 4.8 dictionary meanings, followed by sight words with 3.8 meanings, smell words with 2.5 meanings, taste words with 2.3 meanings, and sound words with 1.7 meanings. Finally, dictionary meaning counts were also reliably affected by the five continuous perceptual strength measures (controlling for Log Frequency, $F(5, 365) = 7.36$, $p < 0.0001$; $R^2 = 0.05$).[6] For three sensory modalities, the regression coefficients were indicated to be reliably different from zero. For touch, there was a positive relationship to semantic complexity (+0.03, $SE = 0.007$, $p < 0.0001$). There were negative relationships for sound (–0.03, $SE = 0.008$, $p = 0.003$) and sight (–0.03, $SE = 0.01$, $p = 0.02$).

How do the three different classes of models (clusters, categorical, continuous) compare against each other? Again, the continuous model had the lowest BIC (–113), followed by the categorical model (–111), followed by the cluster model (–90). Thus, semantic complexity is another linguistic variable for which it helps to know to which sensory modality a word belongs.

15.4. Iconicity

Finally, let us investigate whether form–meaning correspondences differ between the senses. In particular, given the discussion in Chapter 3, we expect a preference for sound concepts to be encoded in an iconic fashion. Many researchers have noted that sound concepts are often onomatopoetic, such as Sweetser (1990, p. 35): "Words for physical sound have most commonly an onomatopoetic origin." And indeed, just looking at the list of sound adjectives from Lynott and Connell (2009), there appear to be many onomatopoetic adjectives, such as *meowing*, *moaning*, *murmuring*, *rustling*, *thudding*, and *thunderous*.

A heightened degree of iconicity appears to characterize deverbal sound adjectives in particular. Adjectives that do not derive from verbs—such as *loud*, *quiet*, and *mute*—do not appear iconic. Perry et al. (2015) and Winter et al. (2017) found that verbs are rated to be more iconic than adjectives in English, so it is plausible that deverbal adjectives may be more iconic as well. Here, we want to establish whether the senses differ in their iconicity, and, if so, whether this depends on a word's association with the verbal domain.

How can iconicity be quantified? One approach is to use native speaker judgments about whether a word is iconic or not. This method was first used for signed languages, including German Sign Language (Grote, 2013, Chapter 3.3) and British Sign Language (Vinson, Cormier, Denmark, Schembri, & Vigliocco, 2008). Perry, Perlman, and Lupyan (2015) were the first to collect iconicity ratings for spoken language, for a set of 592 English and Spanish words. The English part of this dataset was extended by Perry et al. (2018) and Winter et al. (2017), which yielded a total of 3,001 English words normed for iconicity.

Participants rated each word on a scale from –5 ("words that sound like the opposite of what they mean") to +5 ("words that sound like what they mean"). Examples of words with high iconicity ratings are *humming* (+4.47), *click* (+4.46), and *hissing* (+4.46). Examples of words with low iconicity ratings are *miniature*[7] (–1.83), *hamster* (–1.9) and *innocuous* (–1.92). Similar to other constructs in this book, iconicity was thus treated as a continuous quality, with some words being relatively more iconic and some words relatively less (cf. Thompson & Estes, 2011). Iconicity ratings exist for 422 of the 423 adjectives (99%). The ratings are shown for all clusters in Figure 19.

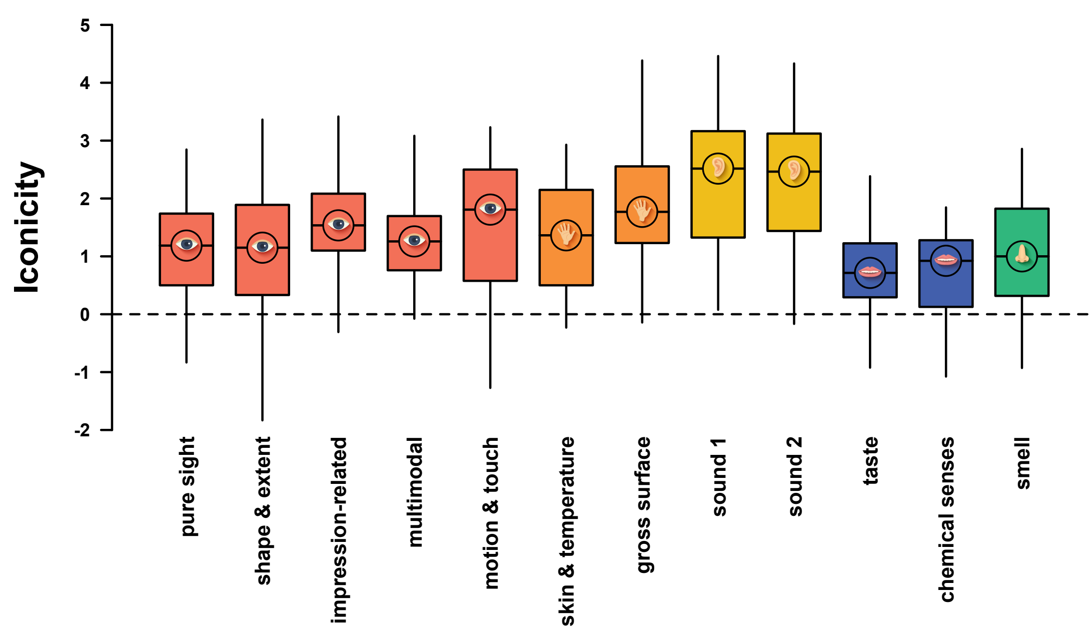


Figure 19. Iconicity ratings by cluster.

As can be seen, the two sound clusters have the highest iconicity values. The clusters associated with taste and smell have the lowest. The sight-related clusters also have relatively low iconicity ratings overall. Within the tactile modality, the "gross surface" cluster is higher in iconicity than the "skin & temperature" cluster. This corresponds to the observation that words involving active touch (i.e., haptic exploration) actually have a very high degree of iconicity in English (see Winter et al., 2017), and in other languages (e.g., Dingemanse, 2011; Dingemanse & Majid, 2012; Essegbey, 2013). There are also many experimental studies showing connections between touch-related concepts and speech sounds (Etzi, Spence, Zampini, & Gallace, 2016; Fryer, Freeman, & Pring, 2014; Moos, Simmons, Simner, & Smith, 2013; Perlman & Cain, 2014).

Modelling iconicity ratings using the cluster model yielded a reliable effect ($F(11, 410) = 8.3$, $p < 0.0001$), which described around 16% of the variance in ratings. The categorical model showed that the five-fold division was also statistically reliable ($F(4, 417) = 21.4$, $p < 0.0001$), describing 16% variance. Iconicity scores were highest for sound (2.3), followed by touch (1.8), sight (1.2), smell (1.0), and taste (0.8). Finally, continuous perceptual strength ratings also had a statistically reliable on iconicity ratings ($F(5, 421) = 17.8$, $p < 0.0001$), describing 17% of the variance. The only two iconicity slopes that were reliably different from zero were sound (+0.27, $SE = 0.04$, $p < 0.0001$) and touch (+0.1, $SE = 0.04$, $p = 0.006$).

The dominant modality model had the lowest BIC (1254), followed by the continuous model (1257) and the cluster model (1290). This is the first dataset explored in this chapter where the categorical model beats the continuous model, although the two models are quite similar in their BICs.

The auditory modality in particular has many deverbal adjectives, which may stem from the fact that sound concepts are more lexically differentiated in the verbal domain (Strik Lievers & Winter, 2018). To assess whether deverbal adjectives are indeed more iconic, etymologies from the Oxford English Dictionary (www.oed.com) were consulted. Of the 68 sound adjectives, 50 traced back to verbs (74%), and only 5 traced back to nouns (7%). The remaining 13 adjectives were adjectives from their first time of attestation. Denominal adjectives were excluded due to their low numbers, allowing for a direct comparison of adjectives and deverbal adjectives. Indeed, deverbal adjectives had higher iconicity ratings (+2.6) than adjectives (+1.2), a reliable difference (Wilcoxon rank sum test, $W = 121$, $p = 0.0005$).[8]

15.5. Conclusions

In this chapter, I used various additional datasets to "make sense" of the sensory adjectives from Lynott and Connell (2009). First, it was shown that English speakers verbalize visual sensations more frequently than sensations associated with the other sensory modalities (Chapter 15.2). This finding suggests that speakers have a greater communicative need to talk about sight compated to touch, sound, taste, and smell. As stated before, this aspect of visual dominance could be due to many different factors; however, the results are at least consistent with the Embodied Lexicon Hypothesis (Chapter 5) because there is independent evidence for visual dominance in perception, as reviewed in Chapter 4. From this perspective, a perceptual asymmetry is reflected in a linguistic asymmetry. The correspondence between type counts (Chapter 12) and token counts (this chapter) suggests that the sensory lexicon is well-adapted to the needs of its speakers: There are more words precisely for those perceptual domains that are more frequently talked about (see Regier et al., 2016; Winter et al., 2018).

This chapter reported two more results. First, touch words were found to be high in semantic complexity, as operationalized through dictionary meanings. This is in line with the observation that touch is frequently used crossmodally, and that touch also frequently metaphorically extends into the mental world (cf. Classen, 1993, Chapter 3; see also Sweetser, 1990). Sound words had particularly low dictionary meaning counts, perhaps because they are more exclusive (Chapter 12) and more restricted in their usage patterns (Chapter 14).

Finally, this chapter showed that knowing about a word's sensory modality is also predictive of form–meaning correspondences. Differences between the senses go all the way down to sound structure. In particular, sound words received high iconicity ratings. Touch words also received relatively high ratings. Taste, smell, and (to some extent) sight were much lower in iconicity. This shows that the sensory vocabulary of English, even when just looking at adjectives, differs with respect to the semiotic strategies identified in Chapter 3. Certain senses prefer certain semiotic strategies.

This set of findings relates to what Dingemanse (2013) calls the "iconicity question": "What are the structural properties of form and meaning such that they afford iconic mappings between the two?" Besides the obvious fact that it may be easier to encode sound with sound due to the overlap in modality (Perlman et al., 2018), the temporal dimension of sound may play a role in the heightened iconicity of words for this sensory modality. Compared to the other modalities, sound is much more dynamic, with sound concepts being about actions and events that are temporally bounded or time-varying (see arguments presented in Strik Lievers & Winter, 2018). This differs from the other modalities, for which there may be a higher number of relatively more static concepts, such as color and shape. The relatively static nature of these non-auditory perceptual features may make it difficult to depict them with speech, which is a dynamic medium after all.

Strik Lievers and Winter (2018) review the evidence for an intrinsic connection between sound and time, which they suggest may lie behind the increased lexical differentiation of sound specifically in the verbal domain, compared to adjectives and nouns. Here, I showed that the among the sound concepts from Lynott and Connell (2009), there are many deverbal adjectives. These deverbal adjectives received higher iconicity ratings than adjectives that were not deverbal. This is consistent with observation that verbs are rated to be more iconic than adjectives (Perry et al., 2015; Winter et al., 2017). This finding additionally suggests that the most "dynamic" adjectives are the most iconic.

Winter et al. (2017) propose that dynamicity may also explain why touch words received relatively high iconicity ratings, since touch often involves the haptic exploration of surfaces, which is a temporally extended process. As stated by Carlson (2010, p. 248), "unless the skin is moving, tactile sensation provides little information about the nature of objects we touch." Although this remains speculative for the time being, it is possible that the dynamicity of touch affords its iconic expression in the dynamic medium of speech. Future research needs to measure the theoretical construct of "dynamicity" directly to provide a quantitative test of these ideas. However, at this stage, the difference between adjectives and deverbal adjectives is already indirect support for the idea that it is precisely the time-varying nature of sound that may make it easier to encode sound concepts in an iconic fashion.[9]

It is worth highlighting that the three variables investigated in this chapter—word frequency, semantic complexity, and iconicity—are each studied in their own right. Thus, an achievement of this chapter is to show that the senses matter when considering these variables. Knowing about a word's sensory modality is partially predictive of various different linguistic patterns.

Finally, let us take a step back and take stock of the evidence for the five senses model. For all datasets considered in this chapter, the continuous model, the categorical model, and the cluster model described linguistic patterns in a statistically reliable fashion. Each model shows that knowing about a word's sensory modality lets us predict different aspects of a word's linguistic properties, ranging all the way from usage frequencies over semantics to sound structure. However, model comparison (using BIC values) showed that treating the senses continuously performed better than treating them categorically in at least two out of three cases. This does not mean that the five senses model or the clusters uncovered in Chapter 13 are wrong in any sense. Instead, it means that for word frequency and dictionary meanings (but not for iconicity), using 12 categorical clusters or five categorical senses does not give additional leverage as opposed to treating perceptual association in a continuous fashion.

The fact that the continuous model consistently performed well has theoretical implications; it suggests that for the linguistic system, discrete labels that are imposed by the analyst do not matter as much. Instead, the continuous degree to which a word is associated with certain modalities appears to be a more important factor. Humans operate within a sensory continuum, and this carries over to sensory language. This is, in fact, another facet of the Embodied Lexicon Hypothesis. To the extent that sensory continuity characterizes actual perception, we should see this characterizes sensory language as well. Humans live in a world characterized by perceptual continuity, and so do their sensory words.

[1] A concern of the analyses presented in this section is that I have only looked at word frequencies from SUBTLEX, a corpus that is, after all, based on movies and TV shows, which may be biased in terms of which sensory words occur. However, similar results are obtained using COCA, and in fact, many commonly used corpora show evidence for visual dominance in word frequencies (Winter et al., 2018).

[2] Word frequencies are a categorical count variable (positive integers) best analyzed using Poisson regression, or negative binomial regression in the presence of overdispersion (see, Zuur, Ieno, Walker, Saveliev, & Smith, 2009). It should be noted that for a similar dataset, Winter et al. (2018) fitted negative binomial models instead of linear regressions on log frequencies. This analysis yielded overall similar results. For the present data, supplementary analyses with negative binomial models are presented in the online scripts (see Chapter 10.7). For all models presented in this chapter, I assessed compliance with the normality and homoskedasticity assumption via residual plots and Q-Q plots. For most models, visual inspection revealed no major problems with these assumptions, except for the continuous semantic complexity model.

[3] Variance inflation factors, however, revealed no major issue with collinearity (all VIF's < 3).

[4] BICs are similar to the more commonly used AICs (Akaike Information Criterion); however, they penalize complex models more strongly.

[5] But see Nakagawa (2012), who argues that the Khoe languages ǂHaba, G|ui, and G‖ana exhibit semantic extensions of taste verbs to touch and hearing.

[6] The slopes of the continuous model were reliably positive for touch (+0.028, SE = 0.007, p < 0.0001). The slopes were reliably negative for sound (–0.03, SE = 0.009, p = 0.003) and, perhaps surprisingly, also for sight (–0.03, SE = 0.01, p = 0.02). If Log Frequency is taken out of the continuous model, there was a reliable positive effect for sight (+0.04, SE = 0.01, p = 0.006), suggesting that the high number of dictionary meanings for sight largely stem from the frequency of sight words. Put differently, there is no "extra" in semantic complexity for being a visual word once frequency is taken into account.

[7] The fact that *miniature* was rated to be one of the least iconic forms is surprising given that the morpheme *mini*– has two high front vowels, which could be taken as an instance of size sound symbolism, especially when contrasted with the form *macro*–. The odd rating for this particular word is probably due to the instructions given to participants. To make sure participants understood the task, several examples of iconic and arbitrary words were presented prior to the task. The demonstration of iconicity emphasized word length and used Hockett's example (1982 [1960], p. 6), who stated that *microorganism* is a long word for a small animal, whereas *whale* is a short word for a small animal.

[8] In an additional analysis, I looked at whether word forms used more frequently as verbs than adjectives are more iconic. For this, the SUBTLEX part-of-speech tags (Brysbaert, New, & Keuleers, 2012) were used. Indeed, there was a reliable difference between verb and adjective usage (W = 195, p = 0.0002) in the predicted direction: The 42 word forms that were used more frequently as verbs had higher iconicity ratings (+1.5) than the 22 word forms that were used more frequently as adjectives (+2.8).

[9] The fact that crosslinguistically, ideophones often encode motion (Dingemanse, 2012, 2018) is consistent with the idea that there is a connection between the time-varying nature of actions and events and a heightened degree of iconicity.

Chapter 16. The evaluative dimension

16.1. Introduction

Wilce (2009, p. 3) says that "nearly every dimension of every language at least potentially encodes emotion" (see also Majid, 2012). Given this, it should come as no surprise that sensory words are also connected to the emotional dimension of language. It is clear that sensory words such as *shiny* and *fragrant* are used often not only to describe perceptual characteristics, but also to perform evaluation (e.g., Lehrer, 1978, 2009). Sensory linguistics needs to consider both perceptual and evaluative meaning, as well as how the two dimensions interact.

This chapter explores how senses as perceptual modalities differ in the degree to which they engage with emotional processes. In particular, taste and smell as purely perceptual modalities are highly emotional. The two modalities have deep cognitive and physiological ties to the human reward system in the brain (e.g., Volkow, Wang, & Baler, 2011; see also Rolls, 2008), and they share close connections with general brain areas for emotional processing (Phillips & Heining, 2002; Rolls, 2008; Royet, Plailly, Delon-Martin, Kareken, & Segebarth, 2000). For example, the olfactory bulb projects directly to the amygdala (Price, 1987; Turner, Mishkin, & Knapp, 1980), a brain area known to be involved in emotional processing (e.g., Halgren, 1992; Richardson, Strange, & Dolan, 2004). Perceiving pleasant or unpleasant odors and tastes is associated with increased blood flow in the amygdala (Zald, Lee, Fluegel, & Pardo, 1998; Zald & Pardo, 1997), more so than for similar visual and auditory stimuli (Royet et al., 2000). Odor classifications are structured along the dimension of pleasantness (Berglund, Berglund, Engen, & Ekman, 1973; Schiffmann, Robinson, & Erickson, 1977; Zarzo, 2008). Behavioral studies show that odors are particularly strong cues for emotionally laden autobiographical memories (Chu & Downes, 2000; Herz, 2004, 2007; Herz & Engen, 1996; Herz & Schooler, 2002; Willander &

Larsson, 2006) or nostalgia (Waskul, Vannini, & Wilson, 2009). Herz (2002) says that "memories evoked by odors are distinguished by their emotional potency, as compared with memories cued by other modalities" (p. 169). Phillips and Heining (2002, p. 204) conclude that taste and smell stimuli are "processed to a significant extent in terms of their emotional content, even if not presented in an emotional context" (see also Yeshurun & Sobel, 2010). Extending from this evidence base, I predict that taste and smell words should also be more emotional. This prediction also follows from the Embodied Lexicon Hypothesis (Chapter 5).

In this chapter, I will begin by reviewing the existing evidence for taste and smell emotionality in language (Chapter 16.2). Then, I will analyze sensory words, first in terms of their overall emotional engagement (Chapter 16.3), then more specifically with respect to positive and negative meanings (Chapter 16.4). In both cases, I will compare the three types of models already introduced in Chapter 15—the categorical model (five senses), the cluster model (twelve subgroups), and the continuous model—to assess the degree to which evaluative language can be predicted by the five senses model.

16.2. Existing linguistic evidence for taste and smell emotionality

The idea that taste and smell language is more emotional, personal, or subjective has repeatedly been expressed within the language sciences (Staniewski, 2017). Lehrer (1978, p. 98, emphasis in original) observed that "*sweet* means 'pleasant' while *sour* and *bitter* connote unpleasantness" (compare Bagli, 2016, 2017). Backhouse (1994, Chapter 1.2) reviews crosslinguistic evidence showing that if a language has only two taste terms, these distinguish between pleasant and unpleasant flavors. Dubois (2000) found that odors are often described with fairly personal language, highlighting the speaker's own involvement rather than an objective description of the odor.

Taste and smell are also embedded in emotionally laden discourse practices. Allan and Burridge (2006, Chapter 8) note how taste and smell are inextricably linked with the culturally loaded domain of food, which gives the corresponding vocabulary special meaning. Backhouse (1994) says that "taste perception is geared towards activities of eating and drinking which, in humans, are pre-eminently culturally channelled activities" (p. 13). An example of this is how people from different cultural groups stereotype, mock, or insult individuals from other groups via their food practices, as when using the term *Krauts* to refer to Germans. Even within a particular country, food and drink items are often subject to heated debate between different cultural subgroups. For example, Germans from Düsseldorf mock people from Cologne for their local beer, Kölsch. In turn, Germans from Cologne mock people from Düsseldorf for their respective local beer, Altbier.[1]

The English taste vocabulary is also connected to the emotionally laden domain of sexual desire: "Both food and bodies *whet the appetite*, *stimulate the juices*, *make the mouth water*, *activate the taste buds*, *excite*, *smell good*, *titillate*, *allure*, *seduce*" (Allan & Burridge, 2006, p. 194, emphasis in original). Similarly, Jurafsky (2014, p. 102) points to the use of sexual words to talk about food, such as when describing a molten chocolate as *an orgasm on a plate*, or marshmallows as *nearly pornographic*. Finally, Velasco-Sacristán and Fuertes-Olivera (2006) talk about the fact that perfume ads are replete with sexual and romantic symbolism, rather than actual smell descriptions.

Describing something using color terms such as *yellow* is fairly neutral in most contexts. Something can be *yellow* without necessarily being attractive or unattractive. However, describing something as *fragrant* or *smelly* is inherently evaluative, which was already observed by Buck (1949) in his dictionary of Indo-European synonyms: “Words for ‘smell’ are apt to carry a strong emotional value, which is felt to a less degree in words for ‘taste’ and hardly at all in those for the other senses” (p. 1022).

There clearly are emotionally valenced terms for the other senses as well; for instance, the word *ugly* describes a negative visual quality. However, for olfaction and gustation, the evaluative component appears to be more obligatory (Levinson & Majid, 2014, p. 411). Evaluative meaning is optional or less pronounced for sight, sound, and touch. Even seemingly neutral expressions involving taste and smell have positive or negative connotations. Krifka (2010) points out that in German, a sentence such as *Der Käse schmeckt* (literally: ‘the cheese tastes’) has a positive connotation, whereas *Der Käse riecht* (‘the cheese smells’) has a negative connotation. This is despite the fact that the two verbs are the most basic perception verbs for the respective modalities, the German equivalents of *to taste* and *to smell* (cf. Dam-Jensen & Zethsen, 2007, p. 1614).

16.3. Absolute valence of sensory words

As discussed in Chapter 5 and Chapter 10, my analysis of evaluative meaning will exclusively focus on the psycholinguistic construct of “emotional valence,” the pleasantness of a word in terms of positive and negative meaning (Warriner et al., 2013; compare Morley & Partington, 2009). I will talk of “evaluative” words as those words that are either very positive, or very negative. As discussed in Chapter 10, it should be kept in mind that by focusing on valence only, the following analyses deliberately ignores the role of specific emotions (such as happiness, joy, disgust, anger and sadness) and additional shades of evaluative meaning (see Bednarek, 2008; Hunston, 2007) to achieve a higher level of abstraction.

The emotional valence of taste and smell words is very obvious in some cases (*stinky*, *smelly*, *fragrant*, *sweet*, *tasty*, *unpalatable*), but not in others (*peachy*, *mushroomy*, *chewy*, *musky*). The subtlety of emotional meaning suggests that an objective measure of valence is required. There are several ways of quantifying the valence of words (Liu, 2012, Chapter 6; Pang & Lee, 2008, Chapter 7; see also Taboada, 2016). The most transparent results are obtained with the native speaker judgments from Warriner et al. (2013), which have been mentioned several times in this book. These authors asked native speakers of English to rate on a scale from 1 to 9 whether a word made them feel "happy, pleased, satisfied, contended, hopeful," or "unhappy, annoyed, unsatisfied, melancholic, despaired, bored." Norms were collected for 13,915 English lemmas. The word with the highest valence value is *vacation* (8.53), followed by *happiness* (8.48) and *happy* (8.47); the word with the lowest value is *pedophile* (1.26), preceded by *rapist* (1.30) and *AIDS* (1.33).

To quantify a word's involvement in evaluative language regardless of whether it is positive or negative, one can compute a word's "absolute valence" (see also Winter, 2016). This was done by *z*-scoring the valence distribution and taking the absolute value of these *z*-scores.[2] What happens when doing this can be demonstrated with the words *sweet* and *moldy*. Whereas *sweet* is about two standard deviations above the average valence (+2.1*z*), *moldy* is about two and a half standard deviations below the average valence (–2.4*z*). When taking the absolute value of –2.4*z*, the negative sign of *moldy* is dropped and the word assumes a value that is close to *sweet* (+2.4*z*.) This corresponds to the fact that both words are highly evaluative, even though one specializes into positive evaluation, and the other specializes into negative evaluation. Thus, the absolute valence measure expresses a word's distance to the mean valence (most neutral words). Whereas low values indicate neutral words, high values are either very positive or very negative. The absolute valence measure ranged from 0 (neutral words with exactly average valence) to 2.7 (highly valenced words that were 2.7 standard deviations above the mean).

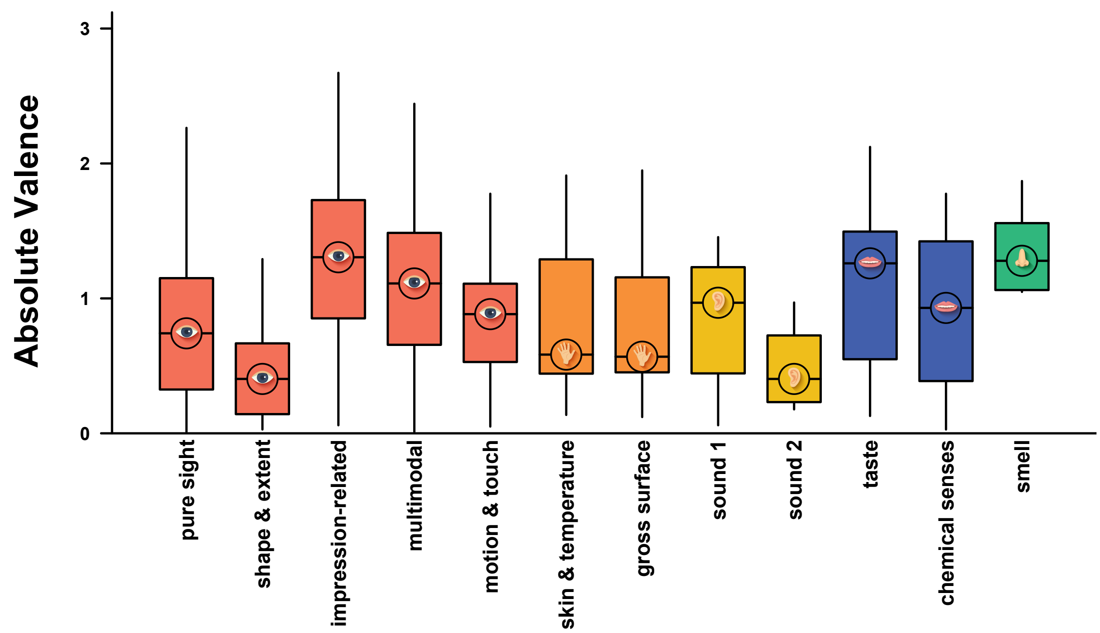


Figure 20. Absolute valence by cluster.

Figure 20 shows the absolute valence measure for the twelve clusters of sensory words. As can be seen, the three clusters associated with taste and smell are especially high on this measure. This reflects the fact that taste and smell words are almost obligatorily positive or negative. The "pure sight" cluster and the "shape & extent" cluster are especially low on this measure. This reflects the fact that words for basic spatial properties have a distinctly neutral feel to them (e.g., *circular*, *compact*, *conical*, *triangular*, *narrow*), as do many color words in the "pure sight" cluster (e.g., *beige*, *blue*, *khaki*, *transparent*). The high absolute valence for the cluster of multisensory "impression-related" words corroborates the suspicion voiced in Chapter 13 that words within this cluster have evaluative functions, including such words as *bloody*, *cute*, *colossal*, *handsome*, *murky*, *ugly*, and *spotless*. It is also interesting that although sound words have overall lower absolute valence, the two sound clusters that were established by the mixture models in Chapter 13 appear to be differentiated along the axis of absolute valence: Words in the "sound 1" cluster are more evaluative, including such negative words as *noisy* and *shrill*.

Fitting the cluster model ($F(11, 246) = 3.7$, $p < 0.0001$; adjusted $R^2 = 0.10$) and the continuous model ($F(5, 257) = 7.9$, $p < 0.0001$; $R^2 = 0.12$) on the absolute valence scores revealed reliable effects in each case.[3] For the continuous analysis, there was only one absolute valence slope that was reliably different from zero, namely the slope of smell (+0.19, $SE = 0.04$, $p < 0.0001$). The positive slope indicates that words with high olfactory strength also had high absolute valence.

The categorical model was not indicated to describe absolute valence in a reliable fashion ($F(4, 253) = 2.05$, $p = 0.09$; $R^2 = 0.02$). On average, smell words had the highest absolute valence ($M = 1.3$, $SD = 0.5$), closely followed by taste words ($M = 1.0$, $SD = 0.6$). Touch words ($M = 0.9$, $SD = 0.5$) and sight words ($M = 0.9$, $SD = 0.6$) had similar absolute valence. Finally, sound words were the most neutral ($M = 0.7$, $SD = 0.5$).

The continuous model performed better (BIC = 463) than the categorical model (487) or the cluster model (494).

However, these results have to be treated with caution because there is relatively little overlap between the Lynott and Connell (2009) norms and the Warriner et al. (2013) norms (only 61%). This is particularly problematic for the cluster model. For example, only seven words in the "sound 2" cluster have absolute valence scores (23% of the 31 words in this cluster), which perhaps makes this result overly dependent on a few isolated words.

16.4. The semantic prosody of sensory words

To circumvent the scarcity of data, one can go beyond looking at the emotional valence of the word itself to the valence of the contexts in which a word occurs. This is related to semantic prosody (Chapter 10.4.2). Snefjella and Kuperman (2016) used the valence norms from Warriner et al. (2013) to compute a measure of context valence based on the average valence of the five content words preceding and following a given head word in the 7 billion token USENET corpus (Shaoul & Westbury, 2013).

In this section, Snefjella and Kuperman's (2016) context valence measure will be used. Because this measure is context-based, words that were not, in fact, normed in the original Warriner et al. (2013) rating study are now associated with a measure of context valence. This means that this measure affords increased descriptive coverage, with 412 of the sensory adjectives being associated with valence data (97%). A measure of absolute valence can also be computed for the context valence, following the procedure described in Chapter 16.3. Thus, what I call "absolute context valence" measures the extent to which a word occurs in contexts that are overall highly evaluative, regardless of whether the evaluation is positive or not. This measure ranged from 0 (words that tend to occur in neutral contexts) to 2.4 (words that tend to occur in highly valenced contexts). Figure 21 shows the absolute context valence for all twelve clusters.

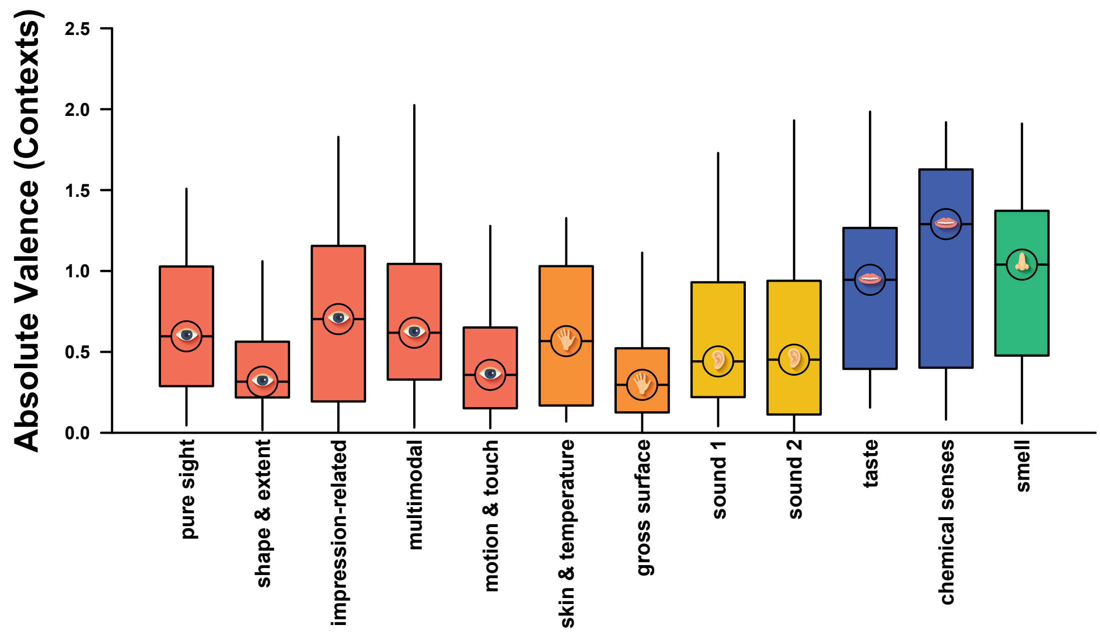


Figure 21. Absolute valence from contexts per cluster.

As can be seen, the clusters associated with taste and smell have the highest absolute context valence. Some of these words did not have particularly high ratings in the Warriner et al. (2013) study. For example, the word *tangerine* in the "taste" cluster was only mildly positive in the rating study (+1.37) but it was the most positive word in terms of context valence (+1.99). Other words in this cluster with high absolute context valence were *lemony*, *delicious*, *tasty*, and *fruity*—all of which are highly positive. Highly valenced words in the "smell" cluster include *fragrant*, *reeking*, *smelly*, *scented*, *stinky*, *perfumed*, and *musky*.

Within the touch and sight modalities, the "impression-related" cluster and the "skin & temperature" cluster have relatively higher context valence. Words with high absolute context valence in the "impression-related" cluster are *lilting*, *cute*, *colossal*, *glamorous*, *clamorous*, *bloody*, *happy*, *mellow*, *radiant*, and *elegant*. Words with high absolute valence in the "skin & temperature" clusters are *stinging*, *tingly*, *cool*, *warm*, *chilly*, *humid*, *tepid*, *cold*, *clammy*, and *fat*. The "shape & extent" cluster is still among the lowest in absolute valence. The "motion & touch" cluster and the "gross surface" cluster also have low context valence. Compared to the valence ratings that were not context based, the "pure sight" cluster has relatively high absolute context valence. The words with the highest absolute context valence in this cluster are *bright*, *silver*, *blonde*, *dazzling*, *brunette*, *blue*, *sunny*, *colorful*, *cloudy*, and *shimmery*.

The cluster model reliably predicted absolute context valence ($F(11, 400) = 5.2$, $p < 0.0001$), which described 10% of the variance in this measure (adjusted R^2). The categorical model reliably predicted absolute context valence as well ($F(4, 407) = 8.7$, $p < 0.0001$) and described 7% of the variance. Taste words assumed the highest values for the absolute context valence measure ($M = 1.0$, $SD = 0.6$), closely followed by smell words ($M = 0.9$, $SD = 0.6$). The other three senses had lower absolute context valence values (sight $M = 0.6$, $SD = 0.5$; sound $M = 0.6$, $SD = 0.5$; touch $M = 0.5$, $SD = 0.4$).

Finally, there was a statistically reliable effect for the continuous model ($F(5, 411) = 16.4$, $p < 0.0001$), which described 14% of the variance. For the continuous analysis, three modality slopes were reliably different from 0, namely sight (–0.06, $SE = 0.03$, $p = 0.01$), touch (–0.07, $SE = 0.02$, $p < 0.001$) and sound (–0.05, $SE = 0.02$, $p = 0.01$). The slopes for taste (+0.03, SE = 0.03, p = 0.22) and smell (+0.03, SE = 0.03, p = 0.34) had positive slopes, but these did not differ reliably from zero. This suggests that in the continuous model for the absolute context valence measure, touch, sound, and sight appear to have a *dispreference* for occurring in emotionally valenced contexts. This dispreference is absent for taste and smell.

Comparisons of these models showed that the continuous model performed best (BIC = 601), followed by the categorical model (627), followed by the cluster model (649).

It should be further noted that modality differences described more variance of the context valence measure from Snefjella and Kuperman (2016) than of the decontextualized valence ratings from Warriner et al. (2013). This suggests that the evaluative differences of sensory words are more directly revealed through looking at word usage in context. In particular, when looking at valence in isolation, taste and smell words do not appear that different from sight, touch, and sound words. When one looks at a context-based measure of evaluative language use, the differences between the chemical senses and the non-chemical senses is more pronounced.

16.5. Positive versus negative valence

The absolute valence measure allowed us to establish that taste and smell words are overall more emotionally valenced. In doing so, we neglected the distinction between positive and negative meaning, which will be analyzed in this section.

Many researchers have noted that languages exhibit negative differentiation with respect to smell (Jurafsky, 2014, p. 96; Rouby & Bensafi, 2002, pp. 148–149): There are more words for malodors (such as body odors and the odors of rotten things) than words for pleasant smells, such as the smell of fresh food. This pattern appears to also characterize languages with larger smell vocabularies than English, such as the Austronesian language Amis, which also exhibits an asymmetry in its smell vocabulary with a relatively larger number of negative words (Lee, 2015). It has been observed by many researchers that even relatively neutral smell words may have negative meaning in context (Alan & Burridge, 2008, Chapter 8; Dam-Jensen & Zethsen, 2007; Krifka, 2010). Lehrer

(2009) observed that the words "*smell* and *odor* are pragmatically negative unless modified by positive adjectives, as in *pleasant odor*, *nice smell*" (p. 249, emphasis in original).

There certainly are a few positive smell adjectives, such as *fragrant* and *aromatic*. However, across the board, there may be a statistical tendency for smell adjectives to be more negative than taste adjectives. This idea can be tested using the raw emotional valence score (without taking the absolute value) from Warriner et al. (2013). For this measure, there was no reliable effect for the cluster model ($F(11, 246) = 1.64$, $p = 0.09$, $R^2 = 0.03$). However, there was a reliable effect for the categorical model ($F(4, 253) = 3.27$, $p = 0.01$, $R^2 = 0.03$). Taste and sight words had the most positive valence (taste $M = 0.2$, $SD = 1.2$; sight $M = 0.2$, $SD = 1.1$). Sound and touch words had slightly negative valence (sound $M = -0.2$, $SD = 0.9$; touch $M = -0.3$, $SD = 1.0$). Smell words were the most negative ($M = -0.5$, $SD = 1.3$).

There also was a reliable effect for the continuous model ($F(5, 257) = 3.6$, $p = 0.004$, $R^2 = 0.05$). There were only two slopes that reliably differed from zero, which was the sight slope (+0.24, $SE = 0.07$, $p = 0.001$) and the touch slope (–0.13, $SE = 0.04$, $p = 0.002$).

The categorical model (796) and the continuous model (797) performed similarly in terms of BIC, and both performed better than the cluster model (829).

It should be noted that the amount of variance described by these models was much lower than that explained by the absolute valence measure. This finding suggests that the differences between sensory words are better characterized in terms of overall valence than in terms of positive versus negative valence. That is, when it comes to sensory language, the different senses differ more with respect to the dimension of neutrality/emotionality than with respect to the dimension positivity/negativity.

How is specialization in positive and negative meanings manifested by the context valence measure by Snefjella and Kuperman (2016)? Here, the cluster model ($F(11, 400) = 3.6$, $p < 0.0001$, $R^2 = 0.07$) and the categorical model ($F(4, 407) = 9.7$, $p < 0.0001$, $R^2 = 0.08$) reliably predicted context valence. Descriptive averages show that taste words occurred in the most positive contexts ($M = 0.7$, $SD = 0.9$), followed by sight ($M = 0.2$, $SD = 0.7$), touch ($M = 0$, $SD = 0.7$) and sound ($M = –0.1$, $SD = 0.8$). Smell words occurred in the most negative contexts ($M = –0.2$, $SD = 1.0$).

There also was a statistically reliable effect for the continuous model ($F(5, 411) = 8.0$, $p < 0.0001$, $R^2 = 0.08$). For the continuous model, there were statistically reliable effects for taste and smell in the predicted direction: Whereas taste had a positive slope (+0.18, $SE = 0.04$, $p < 0.0001$), taste had a negative slope (–0.10, $SE = 0.04$, $p = 0.02$). Thus, the more strongly a word was associated with taste, the more likely it occurred in positive contexts; the opposite is true for smell. The slope for touch also differed reliably from zero (–0.10, $SE = 0.03$, $p = 0.0002$).

Again, it must be noted these valence models described less variance than the absolute valence models reported above. This suggests that even when looking at contexts, differences between the modalities are more pronounced when looking at a sensory word's overall participation in evaluative language, as opposed to a sensory word's participation in specifically positive or negative language. Moreover, just as was found for absolute valence before, the comparison of context valence to the decontextualized valence ratings shows that the evaluative nature of sensory words is better revealed through looking at how a word is actually used in context.

To specifically test the notion that smell is more negative than taste, I performed a planned post-hoc comparison between these two modalities (using the dominant modality classifications). For the valence ratings by Warriner et al.

(2013), there was no reliable difference between taste and smell ($W = 224.5$, $p = 0.09$), although there was a numerical trend in the right direction, with taste words being on average +0.2 standard deviations above the valence mean and smell words being on average –0.5 standard deviations below the mean. The difference between taste and smell words was more reliably revealed through looking at the context valence measure, for which there was a reliable effect ($W = 819$, $p = 0.0005$). On average, the five-word contexts of taste words are +0.7 standard deviations above the valence mean; the contexts of smell words –0.2 standard deviations below the valence mean.

16.6. Conclusions

This chapter showed that the sensory words differ in their evaluative potential. In particular, the analyses presented converging evidence for the idea that the chemical senses, taste and smell, are overall more evaluative. This evidence was found using emotional valence ratings in isolation, but, crucially, it was also found when these ratings were contextualized with a corpus-based measure of context valence (Snefjella & Kuperman, 2016).

It is illustrative to compare the models for absolute valence and absolute context valence discussed in this chapter so far (cluster model, categorical model, continuous model) with an even simpler model that merely distinguishes between the chemical senses (taste and smell) and all other senses. For the absolute valence measure, this model had a BIC of 473, which indicates better performance than the corresponding categorical model (487) and the cluster model (494), but not the continuous model (463). The same ranking appeared for the absolute context valence measure, for which the model that only distinguishes the chemical senses (611) also performed better than the categorical model (627) and the cluster model (649), but not the continuous model (601). This is an interesting observation: Just distinguishing between the chemical senses and the other senses leads to better model performance than incorporating more fine-grained categorical distinctions, such as involved in the five senses folk model. Thus, when it comes to evaluative language, we do not get any extra leverage from distinguishing sight, sound, and touch, and we similarly do not get much leverage out of distinguishing taste and smell from each other.

There are two aspects of the results presented in this chapter that are overall in line with the Embodied Lexicon Hypothesis. Following the correspondence argument (Chapter 5), it should be apparent that there is abundant language-external evidence for the involvement of taste and smell in emotional processes in brain and behavior (see above). This chapter demonstrated this perceptual emotionality corresponds to the emotional qualities of taste and smell language. Both the word-inherent meaning (as gleaned from the decontextualized rating study) as well as the context-based measure suggested a more pronounced emotionality for taste and smell words, in line with the language-external evidence. The second correspondence between language and perception uncovered here harkens back to the discussion of crossmodal connections in Chapter 14, where it was observed that taste and smell pattern together in corpora. The analyses presented in this chapter highlight another way in which taste and smell are quite similar to each other, namely, by virtue of their shared emotional qualities. Taste and smell words not only receive similar modality ratings (Chapters 12 and 13), they also receive similar emotional valence ratings; and taste and smell words not only have similar modality profiles in actual language use (Chapter 14), but also similar evaluative profiles.

The fact that there was some correspondence between valence ratings from a decontextualized rating task (Warriner et al., 2013) and valence in context (taken from Snefjella & Kuperman, 2016) is also relevant to corpus linguistics research. Many researchers have pointed out that the semantic prosody of a word cannot easily be intuited (e.g., Louw, 1993; see also Whitsitt, 2005), but some researchers rightly objected that the mismatch between intuition and contextualized use has never been explicitly tested (Stewart, 2010). The results from this chapter show that contextualized usage can, at least to some extent (and in an admittedly noisy fashion), be intuited. In fact, a correlation test shows that decontextualized valence and context-based valence are fairly strongly correlated with each other (Pearson's $r = 0.65$, $t(256) = 13.6$, $p < 0.0001$), as was shown for a larger dataset in Snefjella and Kuperman (2016). This result is in line with what was found in Chapter 14, which showed that corpus correlations between modality ratings mirrored the correlations that were established in Chapter 13 on the basis of decontextualized modality rating data. Yet again, we find that considerable leverage can be gained from looking at ratings in their own right, and that it is useful to study both contextualized and decontextualized uses together.

The present analyses thus produce two important results for research on semantic prosody. First, they show semantic prosody can be quantified (using data from Snefjella & Kuperman, 2016). Second, the correlation between isolated and contextualized ratings suggests semantic prosody can be intuited across the board. Of course, there are difficult cases to intuit. For example, in the decontextualized rating study, the word *spicy* received a rating that was one standard deviation above the mean of the valence norms. However, it is used in many positive contexts according to the Snefjella and Kuperman (2016) norms, which gives it an even higher context valence score (about two standard deviations above the average context valence). The word *mild* is the opposite. In isolation, it received a somewhat positive rating (+0.7 SDs above the mean).

However, the word’s context valence score is much more negative (–0.8 SDs below the mean). This is presumably because the word is used to modify many negative things, such as *mild recession*, *mild depression*, and *mild fever*. Thus, the mismatch between the isolated and contextualized valence scores yields a quantitative measure of how much intuition fails native speakers for particular words.

There was also statistically weaker evidence for the idea that English taste adjectives are overall more positive than smell adjectives. Why would this be the case? Classen (1993, p. 53) provides the following explanation: “We can choose our food, but we cannot as readily close our noses to bad smells” (see also Krifka, 2010). This would entail that on average, humans are more likely to be exposed to unpleasant smells than to unpleasant tastes. In general, it is to be expected that the cuisines of cultures have evolved to fit the tastes of their communities, and individuals learn throughout their lifetime what they like or do not like in terms of food. As a result, adults very rarely taste things they strongly dislike. Intriguingly, Huisman and Majid (2018) report that food odors were rated as more pleasant than nonfood odors.

It is also possible that the exertion of control over taste itself (being able to choose what to put into one’s mouth)—compared to the lack of control over smell (we cannot choose to not smell)—may further contribute to the perceived negativity of smell, as we generally dislike things that we have less control over (see, e.g., Casasanto & Chrysikou, 2011). These cultural beliefs may further taint smell language. Altogether, there are many potential factors that may explain the relative negativity of smell words compared to the relatively more positive language of taste.

However, despite some negative differentiation for odors and positive differentiation for tastes, both modalities are ultimately associated with both positively and negatively valenced words (e.g., the smell word *fragrant* is positive; *stinky* is not). Given that communicating the distinction between good and bad tastes and smells is quite important (e.g., telling a family member that something tastes moldy), both good and bad words should exist for both sensory modalities.

In sum, this chapter provided another set of comparisons for the three models from Chapter 15 (the continuous model, the categorical model, and the cluster model). Once again, the continuous model outperformed all other models. This reveals another dimension of language use (in this case, evaluative use) better predicted by a graded model of the senses than by a categorical one. As with other linguistic patterns uncovered in this book, evaluative language is characterized by sensory continuity.

---

[1] Kölsch is obviously superior to Alt.

[2] *z*-scoring involves centering the distribution by subtracting the mean and subsequently dividing each value by the standard deviation. The latter step means that the measure is then expressed in standard units: A word with a value of $+1z$, for example, has a valence score that is one standard deviation above the mean; a word with $-1z$ has a valence score that is one standard deviation below the mean. Centering the distribution is necessary to derive the absolute valence measure. The absolute value function makes negative numbers positive. Words with an absolute value close to 0 are neutral words. Words with an absolute value far away from 0 are relatively more evaluative. The extra step of dividing by the standard deviation is not strictly speaking necessary for this measure, but it helps to make the Warriner et al. (2013) scores and the Snefjella and Kuperman (2016) scores more comparable.

[3] For each model presented in this chapter, I assessed compliance with the normality and homoskedasticity assumption by means of residual plots and Q-Q plots. Visual inspection of these plots revealed no major problems with these assumptions.

Chapter 17. Re-evaluating the hierarchy of the senses

17.1. Introduction

The idea that sensory words are used according to a hierarchy of the senses is a compelling one (Chapters 8–9). There have been a number of quantitative studies claiming to find support for the hierarchy (Ronga et al., 2012; Shen, 1997; Strik Lievers, 2015; Ullmann, 1959). However, given the methodological concerns outlined in Chapters 10 and 11, it is necessary to reassess this evidence. In particular: Is there evidence for the hierarchy using the sensory modality classifications from Lynott and Connell (2009)?

Even though I have argued throughout this book that a word's association with the senses is best treated continuously (particularly Chapters 11, 15, &16), this chapter will begin by treating sensory words in a categorical fashion. This allows us to assume common ground with the existing literature on synesthetic metaphors. The categorical assumption is, however, only a temporary one and will be relaxed toward the end of this chapter. In the end, it turns out that the analyses presented here are indeed consistent with the views presented in Chapters 6 through 9; namely, synesthetic metaphors are neither synesthetic nor metaphorical, and there is no monolithic hierarchy of the senses. However, to get there, I first assume the same constraints that characterize this literature.

This chapter proceeds as follows. After asking how one is to interpret corpus data so that it can be counted toward or against the hierarchy (Chapter 17.2), I will present a new analysis that on the surface appears to support the hierarchy of the senses (Chapter 17.3). The point of this analysis is to show that given the methods used throughout this book, it is possible to replicate what has been achieved in the literature on synesthetic metaphor. However, in a second step, I will show that there are problems that have been overlooked (Chapter 17.4). Overcoming these problems leads to alternative analyses that cast serious doubt on the notion of a monolithic hierarchy of the senses. In a final analysis, I will demonstrate the relevance of two linguistic factors in predicting crossmodal language use: namely, the factor of emotional valence and the factor of iconicity (Chapter 17.5). The empirical evidence presented for multiple explanatory mechanisms provides further evidence against the notion of a one-size-fits-all principle. Chapter 17.6 concludes by incorporating the evidence presented in this chapter with the theoretical background on crossmodal language outlined in Chapters 6 through 9.

17.2. What counts as evidence for the hierarchy?

Before engaging with the empirical evidence, one must specify which asymmetries count toward the hierarchy. The most common approach in research on synesthetic metaphors is to create crosstabulations of the senses, as already discussed in Chapter 11 with respect to Ullmann's data from Byron. When using such tables to compare counts of hierarchy-consistent versus hierarchy-inconsistent cases, it is important to specify what counts as hierarchy-consistent or not, especially in the presence of different variants of the hierarchy.

Figure 22 provides a helpful guide. In this matrix, rows indicate sources; columns indicate targets. The blue cells along the diagonal of the matrix are all cases of within-modality uses, such as when a touch-related adjective is used to modify a touch-related noun (e.g., *abrasive contact*). All cells off the diagonal correspond to crossmodal uses. The beige cells in the upper right triangle are those cells that are consistent with what in Chapter 8 I called the "simplified consensus hierarchy" (touch > taste > smell > sight/sound). White cells in the lower right triangle correspond to hierarchy-inconsistent uses of sensory words. If a corpus analysis shows higher word counts in the beige cells than the white cells, the hierarchy of the senses is supported (see Shen, 1997).

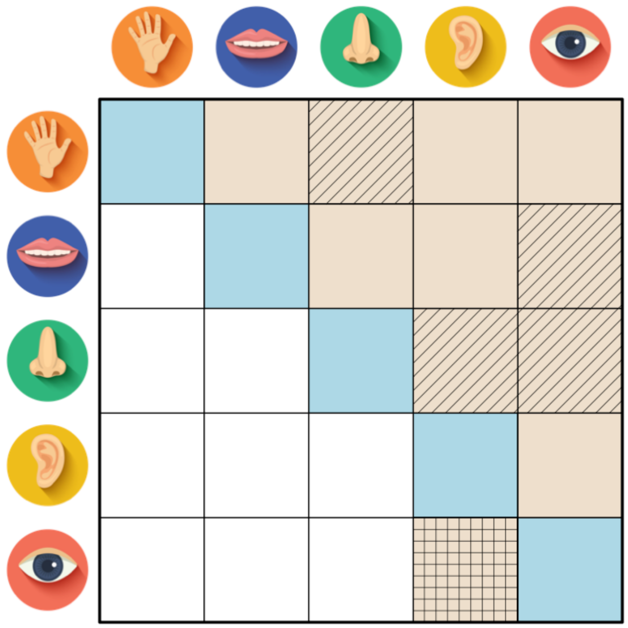


Figure 22. Matrix of source-to-target combinations for all modalities with rows as sources and columns as targets. Diagonal shows within-modality uses; off-diagonal shows crossmodal uses. Beige cells correspond to the simplified consensus hierarchy. The plaid cell is sight-to-sound, which can be interpreted as going against Ullmann's original hierarchy as it is frequently cited (barring the discussion in Chapter 5.22 of his 1959 book). The cells with diagonal stripes are ruled out by the hierarchy of Williams (1976).

The cell with the plaid pattern is the sight-to-sound cell. This cell can be counted toward or against the hierarchy, depending on how Ullmann's original treatment is interpreted (see Shen, 1997, p. 50). The striped cells are mappings that are ruled out by the hierarchy of Williams (1976; I ignore his category of dimension words in this case).

In Chapter 10, I critiqued a table of source-to-target mappings by Ullmann (1945) to exemplify certain methodological concerns. In particular, I argued that one needs to be explicit about what goes into each cell and how sensory words are classified. Ronga et al. (2012), Strik Lievers (2015), and Ronga (2016) circumvent these concerns by preparing a list of sensory words before they approach any corpus data. This way, sensory modality classifications are not made ad hoc, and furthermore, the analysis of crossmodal language can be automated with corpus tools. Here, I will follow these methodological improvements and use the Lynott and Connell (2009) word list to create a crosstabulation in line with Ullmann's original analysis.

In addition, Ronga (2016) reminds us that when studying crossmodal language with a corpus, it is important to distinguish between type and token frequencies. For example, the expression *soft voice* occurred 368 times in COCA (tokens), but it only counts as one adjective–noun pair type. If one only looked at token counts, the results may be biased by a small number of expressions that are highly conventionalized. If all touch-to-sound mappings turned out to be *soft voice*, the evidence for this mapping being part of a hierarchy of the senses would be severely diminished. On the other hand, if there were many different adjective–noun pair types that follow the same tendency, the evidence is more convincing.

One may also consider the number hapax legomena as a measure of the productivity of a particular combination of two senses (see Ronga, 2016). A hapax legomenon is a word (or in this case an adjective–noun pair) that occurs only once in a corpus. Because *soft voice* has multiple instantiations in the Corpus of Contemporary American English (token frequency > 1), it disqualifies as a hapax legomenon. On the other hand, the pair *abrasive warble* would be counted because it occurs exactly once in the COCA data used here. If there were many such hapax legomena that exhibit the same crossmodal combination (e.g., touch-to-sound), then there would be a lot of support for this "mapping" being productive (i.e., it can readily be extended to novel expressions). By comparing word token, word type, and hapax legomena counts, one gets a richer picture of the evidence for the hierarchy of the senses.

17.3. Analysis and results

The following analyses use the dominant modality classifications from Lynott and Connell (2009) together with 219 nouns from Strik Lievers (2015). The reason for choosing these nouns is that they are arguably better sensory nouns than those provided by Lynott and Connell (2013), even though they are not norm-based. As was discussed in Chapters 11 and 12, the nouns from Lynott and Connell (2013) are highly multisensory, and they do not specifically relate to particular senses very strongly, such as the highly abstract words *suspicion*, *rent*, and *comedy*. Moreover, there are no olfactory words in the dataset, and almost all words are dominantly visual. This makes this dataset unsuitable for the present purposes.

The 219 nouns from Strik Lievers (2015) include 133 sound nouns (e.g., *voice*, *whirr*, *rattle*), 49 sight nouns (e.g., *glitter*, *scarlet*, *shadow*), 15 smell nouns (e.g., *perfume*, *stench*, *noseful*), 14 taste nouns (e.g., *savor*, *sapidity*, *flavor*) and 8 touch nouns (e.g., *touch*, *coldness*, *itch*).

The COCA adjective-noun pairs from Chapter 14 were reused for this analysis. As a reminder, this dataset includes about 150,000 adjective-pair types (N = 149,387) that are formed with one of the words from Lynott and Connell (2009). Of the total number of adjective–noun pairs, about 4,500 (N = 4,471) were formed with nouns from Strik Lievers (2015). This list was further paired down, first by excluding words that Strik Lievers (2015) treats as auditory but actually refer to instruments (*lute*, *viola*, *piano*) rather than sound impressions. Such instrument nouns were excluded because they create spurious crossmodal uses—for instance, *black piano* and *red lute* are not audiovisual expressions but literal descriptions of the visual characteristics of instruments.

Following Ronga et al. (2012), 36 dimension words were also excluded.[1] These words are highly multisensory and mostly classified as visual by the native speakers in Lynott and Connell (2009). The dimension words excluded were (in alphabetical order): *big*, *broad*, *bulky*, *chubby*, *colossal*, *compact*, *deep*, *empty*, *enormous*, *fat*, *flat*, *gigantic*, *high*, *hollow*, *huge*, *immense*, *large*, *little*, *long*, *low*, *miniature*, *narrow*, *open*, *petite*, *puny*, *shallow*, *sheer*, *short*, *skinny*, *small*, *steep*, *tall*, *tight*, *tiny*, *uneven*, *wide*. Shape words such as *angular*, *conical*, and *rectangular* were included. Following the native speaker ratings of Lynott and Connell (2009), these were classified as visual.

The following analyses are targeted at assessing the validity of the hierarchy of the senses. Because this hierarchy is about clearly delineated sensory modalities and their relative connections, it is best to work with only those words that are very strongly tied to particular sensory modalities. One advantage of using the modality norms is that quantitative criteria can be used to exclude highly multisensory words (Chapters 11–12). To this end, I analyzed only the 80% most unisensory words (i.e., those words that were above the 20th percentile of exclusivity values). The 20% most multisensory words (such as *strange*, *unripe*, and *strong*) were excluded.

Finally, as one last additional data cleaning step, the highly frequent nouns *look* and *eye* were excluded. This was done because initial inspection of the data revealed that almost all instances of pairs with the word *look* referred to an action, as in *He gave her a sour look*; and almost all instances of *eye* were literal uses describing eyes (such as *blue eye* and *brown eye*) rather than descriptions of sensory impressions.

After all of these exclusions, the remaining dataset contained about 2,600 pair types (*N* = 2,571) and 15,000 tokens (*N* = 14,652). This will form the basis of all analyses presented below. It has to be emphasized that these adjective–noun pairs combine what Werning et al. (2006) call "strong" and "weak" synesthetic metaphors (see also Petersen et al., 2008; Chapter 6.2). That is, cases such as *cold sound* and *cold anger* are lumped together; genuine perceptual and non-perceptual uses of sensory adjectives are not distinguished. This is potentially problematic when comparing the data to other studies on crossmodal language (such as the more carefully constructed datasets by Ronga et al., 2012, and Strik Lievers, 2015). However, to some extent at least, the proof is in the results: Similar patterns to what has been reported in the literature can be obtained using my approach. Moreover, several authors already pointed to the possibility that strong and weak synesthetic metaphors may require the same set of explanations (see Abraham, 1987, p. 179; compare Engstrom, 1946). In essence, the analyses presented in this chapter show how far we can take an approach that does not hand-check each individual crossmodal expression.

Figure 23 shows a crosstabulation of the source–target combinations for the cleaned data. The counts in this matrix are adjective–noun pair *tokens* (not types); that is, these are the raw frequencies of particular source-to-target mappings that are attested in COCA. Because we are looking at tokens, not types, these counts disregard the fact that specific adjective–noun pairs may be overrepresented (e.g., the type *soft voice* is associated with 368 tokens).

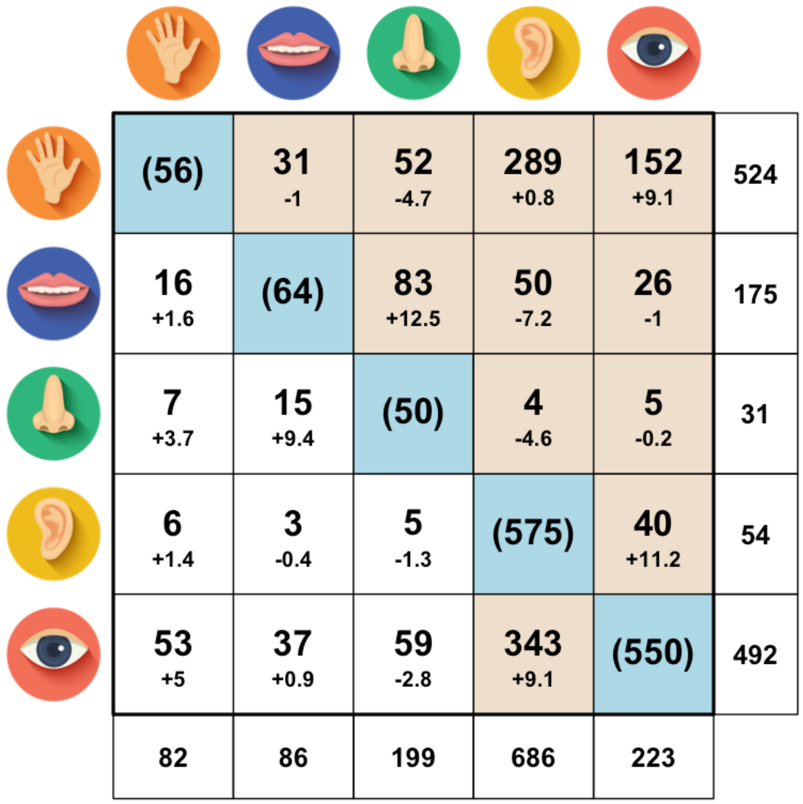
	(288)	56	275	1677	560	2568
	66	(734)	742	288	40	1136
	10	51	(441)	4	6	71
	7	5	10	(4251)	108	130
	416	75	149	1080	(3313)	1720
	499	187	1176	3049	714	

Figure 23. Token counts of sources and targets. Contingency table constructed from the Lynott and Connell (2009) adjectives and the Strik Lievers (2015) nouns. Same-modality cases are bracketed and not included in the row and column marginal frequencies.

The diagonal (blue cells) shows that within-modality uses of sensory adjectives were quite frequent. In this case, they account for about 62% of the tokens. That is, sensory adjectives of a particular modality were most often used to modify nouns from the same modality. In contrast, crossmodal uses were relatively more infrequent.

The row totals in Figure 23 exclude within-modality counts. In terms of raw numbers, adjectives associated with touch (2,568) were used most frequently as sources, followed by adjectives associated with sight (1,720), taste (1,136), sound (130), and smell (71). On the other hand, sound was by far the most frequent target (3,049), followed by smell (1,176), sight (714), touch (499), and taste (187). These numbers can be used to calculate source-target ratios (i.e., dividing the row total by the column total). A ratio larger than one indicates that a given modality is used more as a source than as a target; a ratio smaller than one indicates that a modality is used more as a target than as a source. These numbers indicate that taste (6.07) and touch (5.15) were used much more as sources than as targets—namely, about 5–6 times as much. The same was the case for sight (2.41), although to a much lesser extent. Finally, smell (0.06) and sound (0.04) occurred more frequently as targets rather than as sources.

Following Shen (1997) and others, I compared the overall count of hierarchy-consistent cases (beige cells) to the overall count of hierarchy-inconsistent cases (white cells). Overall, 4,836 out of 5,625 tokens were in those cells that are consistent with the simplified consensus hierarchy, which is 86% of all tokens. A simple binomial test indicates this to be reliably different from chance ($p < 0.0001$, using a chance baseline of 55% for 11 out of 20 cells). If the sight-to-sound cell is excluded from counting toward the hierarchy (strictly following Ullmann's original hierarchy), the percentage of hierarchy-consistent cases drops to 67%. A binomial test indicates this to be reliably different from chance ($p < 0.0001$, using a 50% baseline for 10 out of 20 cells). For the hierarchy of Williams (1976), the match is 80%, reliably different from chance ($p < 0.0001$, using a 35% baseline for 7 out of 20 cells). These binomial tests have to be interpreted with care as the same adjective–noun pair contributes multiple data points, which violates the independence assumption of this statistical test. This concern can be alleviated by analyzing adjective–noun pair *types*, rather than tokens. These types are shown in Figure 24.



(56)	31 -1	52 -4.7	289 +0.8	152 +9.1	524
16 +1.6	(64)	83 +12.5	50 -7.2	26 -1	175
7 +3.7	15 +9.4	(50)	4 -4.6	5 -0.2	31
6 +1.4	3 -0.4	5 -1.3	(575)	40 +11.2	54
53 +5	37 +0.9	59 -2.8	343 +9.1	(550)	492
82	86	199	686	223	

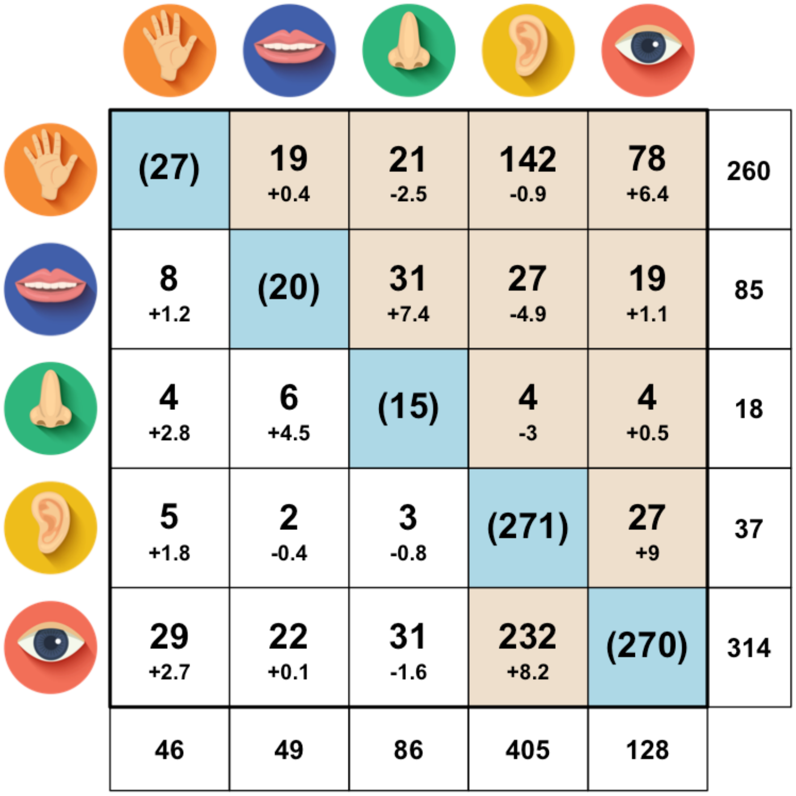
Figure 24. Type counts of sources and targets. Values below each count are adjusted standardized Pearson residuals, calculated without taking the diagonal into account.

The small numbers below each type count in Figure 24 are standardized Pearson residuals (calculated without taking the diagonal into account). Positive residuals indicate overrepresentation; negative numbers indicate under-representation. As a rule of thumb, values larger than |2| can be interpreted to indicate reliable over- or under-representation (Levshina, 2015, pp. 220–221).

Again, inspecting the table shows that within-modality uses of sensory adjectives were relatively frequent, accounting for about 50% of the data. Next, looking at the row totals (which exclude within-modality counts) reveals that touch adjectives were used most frequently to talk about the other senses (524); that is, touch was the most prolific source domain. This was followed by sight (492), taste (175), sound (54), and smell (31), respectively. Looking at the column totals reveals that sound nouns were most often modified by words from the other senses (686); that is, sound was the most frequent target domain. This was followed by sight (223), smell (199), taste (86), and touch (82), respectively. As was done for token counts, the source and target frequencies can be brought into correspondence with each other via source–target ratios, which was highest for touch (6.39), followed by sight (2.21), taste (2.03), smell (0.16), and sound (0.08).

A comparison of hierarchy-consistent cases to hierarchy-inconsistent cases reveals that 84% of the type counts fit the simplified consensus hierarchy (binomial test: $p < 0.0001$, using a chance baseline of 55% for testing 11 out of 20 cells). After excluding the sight-to-sound cell, there are only 57% hierarchy-consistent cases, although a binomial test indicates this to be still reliably different from chance ($p < 0.0001$, using a 50% baseline for 10 out of 20 cells). For the hierarchy of Williams (1976), the match is 77%, reliably different from chance ($p < 0.0001$, using a 35% baseline for 7 out of 20 cells).

Finally, let us look at counts of hapax legomena (see Ronga, 2016), as shown in Figure 25.



(27)	19 +0.4	21 -2.5	142 -0.9	78 +6.4	260
8 +1.2	(20)	31 +7.4	27 -4.9	19 +1.1	85
4 +2.8	6 +4.5	(15)	4 -3	4 +0.5	18
5 +1.8	2 -0.4	3 -0.8	(271)	27 +9	37
29 +2.7	22 +0.1	31 -1.6	232 +8.2	(270)	314
46	49	86	405	128	

Figure 25. Hapax counts of sources and targets. Values below each count are adjusted standardized Pearson residuals, calculated without taking the diagonal into account.

These counts corroborate the results of the type counts. Source–target ratios are highest for touch (5.65), followed by sight (2.45), taste (1.73), smell (0.21), and sound (0.09). As was the case for type and token counts, only smell and sound occurred more often as targets than as sources.

Let us draw some interim conclusions. I have presented new evidence for the hierarchy of the senses. In line with prior reports, touch was the most frequent source domain and sound was the most frequent target domain. However, smell was another frequent target, as well as an infrequent source.

This is consistent with what Ullmann (1959) and Tsur (2008) noticed. It is also consistent with the fact that smell is not a possible source in the hierarchy of Williams (1976; see also Chapter 8). The fact that smell is a frequent target in crossmodal expressions furthermore fits the view that smell is lexically under-differentiated, which may necessitate the use of metaphor.

It has to be emphasized that the percentage of hierarchy-consistent cases reported here is surprisingly similar to what has been reported in the literature. Shen (1997) reported a match of 91% for his Hebrew corpus if sight-to-sound was included and 73% if sight-to-sound was excluded (see also Whitney, 1952). Strik Lievers (2015) reports perhaps the lowest figures, with 62% consistent cases for English and 74% for Italian.[2] It is also noteworthy that the highest match was observed for token counts, with lower percentages observed for type and hapax counts. This suggests that stronger evidence for the hierarchy is obtained if highly conventionalized adjective–noun pairs are taken into account.

Of the three different versions of the hierarchy analyzed here, the simplified consensus hierarchy (as embodied, for example, by Shen's work) consistently had the highest percentage of hierarchy-consistent cases. However, the hierarchy by Williams (1976) was a close second, and one has to acknowledge that this hierarchy is also more parsimonious (it reaches a high percentage of hierarchy-consistent cases while at the same time positing fewer connections).

Interestingly, similar to what was found here, Strik Lievers (2015) also reports that a comparably high number of sight adjectives acted as sources in her data, more so than what is perhaps expected based on past descriptions of the hierarchy, which put sight at the top along with sound. The high number of sight words used as sources may be due to the fact that her dataset, like the dataset considered here, has a large proportion of sight adjectives. To control for this, the number of adjectives needs to be compared against a baseline. Moreover, it must be emphasized that the binomial tests conducted so far have violated the independence assumption because the same adjective can recur across multiple different adjective–noun pairs (i.e., there are repeated measures for adjectives). To circumvent these concerns, the following analysis uses the total set of adjectives from Lynott and Connell (2009) as a baseline. I then look at how many of the adjectives in this list are used for each crossmodal combination. In essence, this analysis is treating the Lynott and Connell (2009) word list as the pool of words that crossmodal expressions can be drawn from. Figure 26 shows the percentage of adjectives used crossmodally. In this matrix, a value close to 100% means that nearly all of the adjectives from that sensory modality were used in the corpus; values close to 0% mean that adjectives from that modality were underutilized.

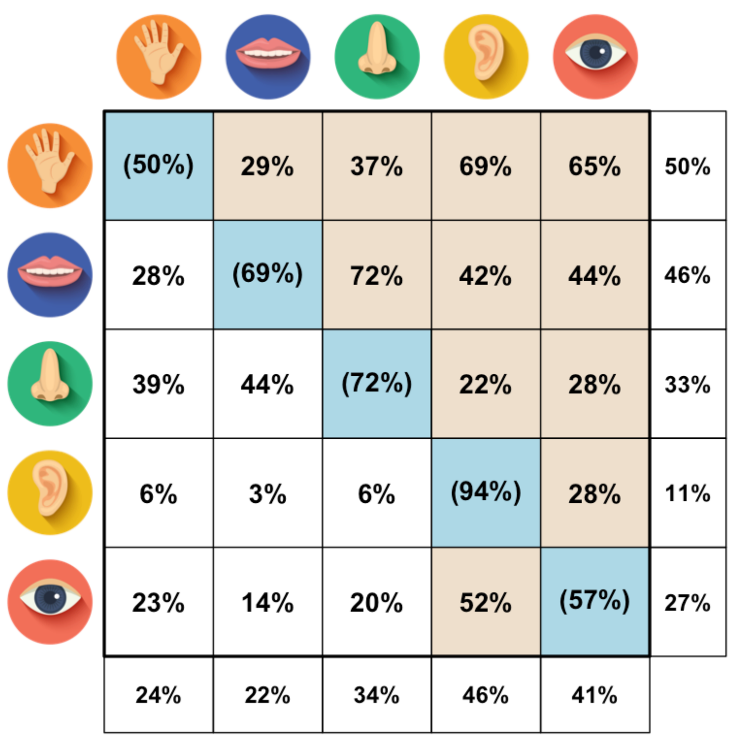


Figure 26. Percent of adjectives used (out of the adjective list from Lynott and Connell, 2009). Same-modality cases are bracketed and not included in the row and column averages.

Let us first look at the row averages. These show that on average, about 50% of the touch words from the Lynott and Connell (2009) dataset occurred in descriptions with targets from the other sensory modalities. In terms of source percentages, the next-highest modality was taste (46%), followed by smell (33%), sight (27%), and sound (11%). Column averages reveal that taste (22%) and touch (24%) attracted less distinct adjectives from the other modalities. Smell (34%), sight (41%), and sound (46%) attracted comparably more adjectives in crossmodal usage. On average, hierarchy-consistent cases (using the consensus hierarchy) had a higher percentage of adjectives used (44%) than hierarchy-inconsistent cases (20%; Wilcoxon rank sum test, $W = 84$, $p = 0.01$).

Source-target ratios of these percentages reveal that touch (2.08) and taste (2.06) were used about twice as likely as sources, followed by smell (0.99), which was equally likely to be used as source and target. Sound (0.23) and sight (0.66) were less likely targets than sources. Thus, once controlling for lexical differentiation (in particular, the overrepresentation of sight words in English), the data resembles the hierarchy of the senses even more closely. With this analysis, sight appears as a more frequent target, next to sound.

17.4. Deconstructing the hierarchy of the senses

The evidence presented so far looks remarkably consistent with what has been reported in the literature on synesthetic metaphors, even though I have used a different dataset (COCA), a different way of classifying the senses (Chapter 10), and my analysis did not distinguish between strong and weak synesthetic metaphors. On the surface at least, it looks as if there is a lot of independent support for the hierarchy of the senses, especially if we take this study together with already existing empirical studies, in particular the rigorous investigations conducted by Ronga et al. (2012), Strik Lievers (2015), and Ronga (2016).

However, a closer look at the data suggests that the hierarchy may not be a good explanatory model. It should be emphasized that the following critical arguments are not targeted at the hierarchy of the senses as a descriptive generalization. Instead, my arguments are targeted at the idea that there is one unifying principle that explains all of the observed asymmetries (i.e., the idea that there is a *monolithic* hierarchy of the senses), such as is the case with Shen's directionality principle (see Chapter 9.2.1). In essence, my argument will be that the observed descriptive patterns are inconsistent with a one-size-fits-all principle.

First, it is illustrative to have another look at the crosstabulations shown in Figures 23–26 to investigate which specific cells are overrepresented. Ignoring within-modality cases, the largest number of tokens (Figure 23) was obtained for the touch-to-sound cell (1,677). The second largest number of tokens was obtained for the sight-to-sound cell (1,080). The third largest number was obtained for the taste-to-smell cell (742). These three cells alone account for over 62% of the total number of tokens (out of 5,625 tokens that constitute crossmodal uses, excluding within-modality counts). Moreover, these three cells alone account for about two-thirds of the hierarchy-consistent cases (72%, simplified consensus hierarchy). Performing such computations for type and hapax counts paints a similar picture. For both of these modalities, the sight-to-sound, touch-to-sound, and touch-to-sight cells were the three cells with the highest frequencies. These three cells accounted for 61% (type counts) or 62% (hapax counts) of the total number of crossmodal uses, as well as for 73% and 75% of the total number of hierarchy-consistent cases (simplified consensus hierarchy).

Thus, about two-thirds of the hierarchy-consistent cases are accounted for by only three cells out of the eleven cells that are generally treated as being part of the simplified consensus hierarchy. This means that the other eight cells don't do as much "work" in supporting the hierarchy of the senses. It is also noteworthy that except for the case of sight-to-sound, the cells with the highest frequencies were among those particular modality combinations that were argued to be perceptually associated with each other (see Chapter 14; see also arguments in Cacciari, 2008, p. 436; Ronga, 2016; Ronga et al., 2012). This is particularly the case when looking at the standardized residuals, which assumed large positive values (overrepresentation) for the touch-to-sight cell and the taste-to-smell cell—precisely those domains that I have argued are perceptually associated with each other (Chapter 14). With this in mind, the evidence presented in the crosstabulations of this chapter actually looks more consistent with the Embodied Lexicon Hypothesis than with the hierarchy of the senses. The evidence is furthermore consistent with the theoretical account presented in Chapter 7, where I have argued that many crossmodal uses of sensory words are driven by the fact that these words have broad referential meaning encompassing multiple perceptually related modalities.

Moreover, the very fact that the standardized residuals are all over the place—and quite different even between different cells that are counted toward the hierarchy—casts doubt on the idea of a monolithic hierarchy. In Chapter 9, I introduced the distinction between local and global explanatory accounts of the hierarchy. To merely count the proportion of hierarchy-consistent versus hierarchy-inconsistent cases neglects the fact that particular mappings are over- or underrepresented, which would speak to a local rather than global account.[3]

Moreover, across different studies, different crossmodal connections are strong or weak. What appears as hierarchy-consistent in one study may be substantiated by quite different sensory mappings than what appears as hierarchy-consistent in another study. This is the danger of using a gross summary measure, such as the percentage of upwards-transfers. The specific crossmodal connections may not actually support the bigger picture of a monolithic hierarchy, especially if some of them disproportionately contribute to the count of hierarchy-consistent cases.

It should be noted, however, that in this dataset, there is indeed a strong link between sight and sound, as already discussed by Ullmann (1959). The standardized residuals for the sight-to-sound cell and the sound-to-sight cell are both quite high. This specific mapping warrants further explanation. The experimental findings by Marks (1974, 1975, 1982a), already briefly discussed in Chapter 7, may explain this particular crossmodal connection. Marks found that non-synesthetes consistently associate brightness with loudness and pitch. This particular crossmodal correspondence is presumably widely shared in the population and may motivate this language use. However, if this crossmodal correspondence would lie behind the high frequency of audiovisual language, this would be even more evidence for a local as opposed to a global explanatory account. Specifically, Marks' research on sound/sight matching is something that is *specific to these two modalities*. Thus, invoking this as an explanation for the observed asymmetries between sight and sound would lead us further away from a monolithic account of the hierarchy.

Finally, the present data supports the hierarchy of the senses even less if we take the differential ineffability of the senses into account. In Chapter 4, I explored the idea that for certain senses, there are more words than for other senses (see Levinson & Majid, 2014), which was supported by empirical evidence in Chapter 12. When looking at crosstabulations such as shown in Figures 24 through 26, one needs to ask the question: Given the imbalances present in the sensory vocabulary of English, what asymmetries in source-to-target mappings are expected even without invoking any extra principles that pertain specifically to crossmodal expressions? That is, could it be that the asymmetries in the composition of the sensory vocabulary alone play a role in what crossmodal expressions are frequently observed? This was indeed suggested by Strik Lievers (2015), and another test of this idea will be provided here.

In the following analysis, I combined all adjectives from Lynott and Connell (2009) with all nouns from Strik Lievers (2015), which yielded about 80,000 (N = 81,639) adjective–noun pairs, including about 60,000 (N = 62,612) crossmodal combinations. For this set of adjectives and this set of nouns, these adjective–noun pairs exhaust all possible combinations. Then, within this set of crossmodal combinations, I counted the number of hierarchy-consistent cases (type counts). This turns out to be 82% (simplified consensus hierarchy), which was found to be statistically reliable by a binomial test ($p < 0.0001$, with a 55% chance baseline for 11 out of 20 cells). This number is eerily similar to what we observed above on the basis of actual corpus data, where the figure was 84%. A look at source-target ratios reveals the following ranking: touch (4.59) > sight (2.76) > taste (1.87) > smell (0.78) > sound (0.15). This ranking also closely corresponds to what was observed in the corpus data.

This simple computation thus produces a striking result: It shows that even if speakers were to randomly combine adjectives and nouns from the two datasets considered here, evidence consistent with the hierarchy would arise. Therefore, imbalances that are already present in the sensory vocabulary can create patterns that an analyst would interpret to be in line with the hierarchy of the senses. If this is the case, the underlying mechanism is not a hierarchy specific to crossmodal language, but, in fact, asymmetries that characterize the composition of the sensory lexicon in general. This shows that it is important to consider crossmodal language in the broader context of the sensory vocabulary. When this is done, there may not be any extra explanations necessary to account for the observed asymmetries.

The next section presents further evidence against a monolithic hierarchy by demonstrating the influence of two other factors, emotional valence and iconicity (see Chapter 9). This is the first step towards an analysis that explores the role of multiple explanatory constructs in a conjoined fashion. If it is possible to show that multiple factors play a role in accounting for crossmodal language use, a one-size-fits all principle appears less likely.

17.5. Emotional valence and iconicity predict metaphor choice

The evidence presented so far clearly shows that there are asymmetries with respect to which adjective–noun pair combinations are more frequent, even if these asymmetries may be driven by the general composition of the sensory lexicon. However, these asymmetries in and of themselves have nothing to say about what explains the observed patterns, as discussed in Chapter 9. As happened so often in the history of linguistics, the observed descriptive regularity of a hierarchy of the senses was directly interpreted as being a governing principle itself—that is, a ranking seen in the data was taken to reflect a hierarchy of the senses that explains the observed asymmetries. This is circular reasoning (Dąbrowska, 2016a; Gibbs, 2007). To get around this circularity, external datasets are needed. Merely analyzing frequencies of crossmodal uses does not allow strong inferences on specific explanatory accounts. Instead, one needs to correlate the frequency of adjective–noun pairs with other measures that relate to specific explanatory constructs.

In this section, I will consider the role of two linguistic factors that have been hypothesized to drive crossmodal language use: valence and iconicity. As was discussed in Chapter 9, several authors have noted that many of the crossmodal uses of sensory adjectives appear to have evaluative qualities. This predicts that relatively more evaluative sensory words are more frequently used in a crossmodal fashion, compared to relatively more neutral sensory words. Accordingly, the adjective–noun pair *sweet music* is predicted to be more frequent than *palatable music*, given that *palatable* is more neutral than *sweet* (as per the valence norms of Warriner et al., 2013). In addition, Chapter 9 introduced the possibility that iconicity may restrict the domain of application of sensory words. Thus, iconic words are predicted to be used *less* in crossmodal expressions—that is, *squealing color* should be less frequent than *loud color*.

To test these ideas, the valence norms from Chapter 16 and the iconicity norms from Chapter 15 will be used. For the first analysis, the COCA data used to construct the crosstabulations seen in Figures 23–26 was coded for whether an adjective–noun pair indicates a crossmodal use (off the diagonal of the crosstabulations) or a within-modality use (on the diagonal). This binary measure was then subjected to a mixed logistic regression analysis with the fixed effect "Absolute valence" (lower values = more neutral; higher values = more evaluative; see Chapter 16). The model also included random intercepts for adjectives, as well as random intercepts for nouns and by-noun varying random slopes for the effect of absolute valence.

The model estimated a positive relationship between absolute valence and crossmodal use (logit estimate: +0.72, SE = 0.32, Wald's z = 2.25, p = 0.03), as shown in Figure 27a. For each increase in absolute valence by one raw valence point, the odds of being used crossmodally increased 2.05 to 1. A similar analysis with the absolute valence derived from contexts (from Snefjella & Kuperman, 2016) revealed no reliable effect (–0.26, $SE = 0.33$, $z = –0.78$, $p = 0.44$).

The analysis was repeated for iconicity, with a separate model where the absolute valence predictor was exchanged for the iconicity ratings used in Chapter 15. There was a negative relationship between iconicity and crossmodal use (–0.50, $SE = 0.15$, $z = –3.30$, $p = 0.001$), as shown in Figure 27b. For each increase in iconicity by one raw iconicity rating point, the odds of observing a *within-modal* use increased 1.65 to 1.[4]

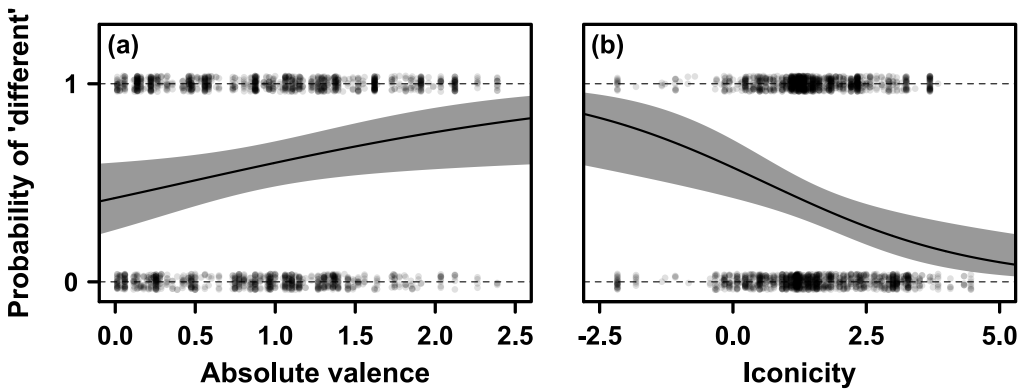


Figure 27. Probability of an adjective being used crossmodally as a function of the adjective's (a) absolute valence and (b) iconicity. Each data point (with vertical scatter for visibility) indicates one adjective–noun pair; shaded areas indicate 95% confidence region; data from 2,571 adjective–noun pairs

The analyses so far were constrained to a categorical classification of the senses—a constraint which Chapter 7 has argued against, and a constraint that is furthermore inconsistent with the evidence presented in Chapters 15 and 16. Now, we will relax this methodological constraint and use the continuous measure of modality association that is afforded by the Lynott and Connell (2009) norms. In Chapter 14, I used cosine similarity as a general measure of modality affinity. Even though the results in Chapter 14 were not discussed this way, they actually do speak to the topic of synesthetic metaphor. Namely, if the cosine similarity between an adjective and a noun is high, then this indicates a within-modality use. If the cosine is low, this indicates a crossmodal use. In fact, the cosine similarity can be interpreted as a continuous measure of metaphoricity, with lower cosines (such as *sweet music*) having a more "metaphoric" feel than words with higher cosines (such as *abrasive contact*), which appear relatively more literal. Is this measure of metaphoricity (as operationalized through modality fit) predicted by valence and iconicity? In line with the theory outlined here (and in Chapter 9), it is expected that relatively more evaluative adjectives occur in adjective–noun pairs with lower cosines (more crossmodal). In contrast, it is expected that relatively more iconic adjectives occur in adjective–noun pairs with higher cosines (less crossmodal).

Mirroring the analyses just performed on categorical sense classifications, cosines were regressed onto absolute valence, absolute context valence, and iconicity. Linear mixed effects models had the same fixed and random effects structure as before. Models were estimated with maximum likelihood and *p*-values were computed with likelihood ratio tests. There was only a numerical trend for absolute valence (–0.03, $SE = 0.014$, $\chi^2(1) = 3.72$, $p = 0.054$). There was, however, a statistically reliable pattern of absolute context valence (–0.08, $SE = 0.01$, $\chi^2 (1) = 38.57$, $p < 0.001$), as shown in Figure 28a. Both patterns were as predicted: Words with higher absolute valence—that is, less neutral words—were more likely to be used crossmodally.

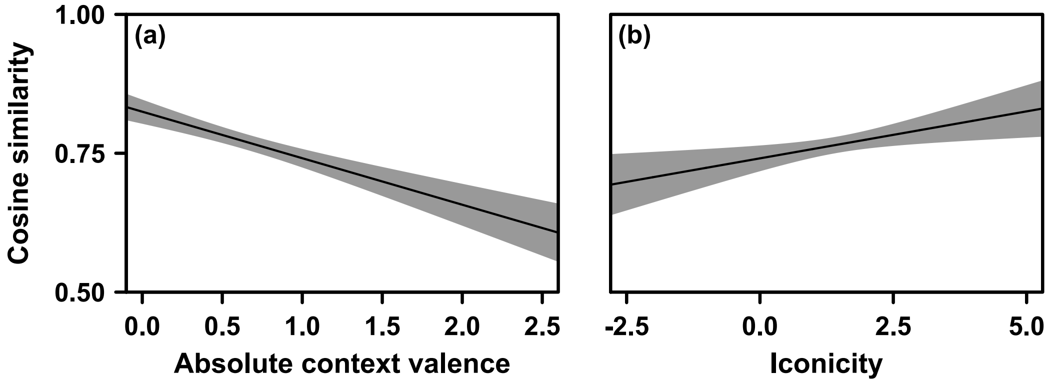


Figure 28. Cosine similarity as a function of the adjective's (a) absolute valence and (b) iconicity. Shaded areas indicate 95% confidence region.

Furthermore, there was a positive correlation between iconicity and cosines: Relatively more iconic adjectives were more likely to occur in expressions where the adjective assumes a similar modality to the noun (+0.17, $SE = 0.006$, $\chi^2(1) = 6.93$, $p = 0.008$), as shown in Figure 28b.

The evidence presented in this section shows at least two factors contribute to crossmodal uses of sensory adjectives: valence and iconicity. Adjectives that were relatively more valenced (less neutral) were *more* likely to be used crossmodally; adjectives that were relatively more iconic were *less* likely to be used crossmodally. It should be emphasized that the two analyses presented here (of the categorical data and of the cosines) share a common denominator, the sensory adjectives of Lynott & Connell (2009). However, other than involving the same adjectives, the two analyses are at least partially independent because they involve different noun sets. In the case of the categorical data, I used the nouns from Strik Lievers (2015). In the case of the cosine data, I used the nouns from Lynott and Connell (2013). The fact that the two analyses converge on the same set of results despite differences in data and method needs to be highlighted. Altogether, this data suggests that it is fruitful to step away from the notion of a monolithic hierarchy of the senses. Instead, we should think about what linguistic factors predict crossmodal uses of sensory words.

17.6. Conclusions

This chapter began by presenting new evidence for the hierarchy of the senses. Using different methods, this chapter obtained results that were highly similar to what has been reported by other empirical studies, such as Ullmann (1959), Shen (1997), Ronga et al. (2012), Strik Lievers (2015), and Ronga (2016). However, I then showed that even though *descriptively* there were clear asymmetries between the senses, this did not provide evidence for a monolithic hierarchy. In particular, specific crossmodal connections were overrepresented, namely those senses that are tightly connected in our everyday lives and in our perceptual systems. From this perspective, the evidence presented here is very much in line with the evidence for modality affinity obtained in Chapter 14. The evidence presented here is also consistent with the Embodied Lexicon Hypothesis (Chapter 5) since yet again, I found that perceptually associated senses are also linguistically associated.

Furthermore, it must be emphasized that there was a large number of within-modality mappings. These have been ignored in previous studies of synesthetic metaphors, but in the present data, they constituted at least half of the uses of sensory words. Thus, one must view the prevalence of crossmodal language within the broader context: In line with the concept of modality affinity (Chapter 14), it seems as if sensory words like to attach to words of the same modality, or to words associated with related modalities. What others call “synesthetic metaphors” is not the dominant pattern in perceptual language.

Perhaps the biggest concern for a notion of the hierarchy was the fact that exhaustively combining the sensory words from the respective adjective and noun word lists created a pattern of asymmetry that looked remarkably like what was seen in the corpus data and what has been reported in other studies. This suggests that when doing studies on crossmodal uses of sensory adjectives, one has to think about the composition of the sensory lexicon. Asymmetries that characterize the perceptual vocabulary of English can create patterns that look like what others have treated as the hierarchy of the senses. This connects the findings presented here with the empirical study of dominance relations discussed in Chapter 12 and more generally, this shows that the concept of the "differential ineffability of the senses" plays a role crossmodal language. Importantly then, the hierarchy of the senses is not a hierarchy that governs specific crossmodal uses (e.g., ruling out sound-to-touch etc.), but a descriptive generalization that at least in part may result from the way the English lexicon is composed. This makes it unnecessary to evoke cognitive constructs such as accessibility (see Chapter 9) to explain the empirically observed asymmetries, as is done by Shen and others (Shen, 1996, 1997, 2008; Shen & Aisenman, 2008; Shen & Cohen, 1998; Shen & Gadir, 2009; Shen & Gil, 2007).

Any notion that is grounded in one monolithic explanatory principle that actively governs specific "mappings" neglects the fact that there are already asymmetries in how the sensory vocabulary is composed, and it furthermore neglects the fact that certain hierarchy-consistent "mappings" are more frequent than others. In addition, I empirically demonstrated that there is more than just one mechanism at play by providing two partially independent analyses demonstrating the relevance of valence and iconicity. As far as I am aware, this is the first direct test of an explanatory account, rather than relying on verbal arguments. It is this kind of empirical evidence that needs to be incorporated into the debate about what explains the hierarchy of the senses.

Each of these two explanatory factors is interesting in its own right. The idea that emotional valence predicts crossmodal uses of sensory adjectives is in line with the proposal that synesthetic metaphors often have affective qualities (Chapter 7.5.2, Chapter 9.3.2). It seems that when speakers use language primarily to evaluate, the referential fit does not have to be as close. As was discussed in Chapter 7.5.2, speakers choose particular words for multiple reasons. The denotational fit (i.e., whether a word identifies a referent and its characteristics precisely) is only one of those reasons. Oftentimes speakers also choose words for reasons of evaluation and the expression of affective content. When doing so, a speaker may be more loose with the denotational fit. Thus, both descriptive and evaluative factors co-determine word choice, but when one factor is foregrounded, the other may be less important. Thus, the evidence presented here is in line with the notion of a trade-off between emotional and perceptual meaning, which is also suggested by the work of Vigliocco et al. (2009) and Kousta et al. (2011).

The idea that iconicity is negatively associated with crossmodal use also warrants further discussion. As was discussed in Chapter 9, there is existing evidence from research on signed languages that the iconicity of signs may block certain metaphorical uses (Emmorey, 2014; Meir, 2010). As was discussed in Chapter 3, iconicity communicates perceptual content very vividly. For sensory words, this means that iconic words are very strongly connected to their associated sensory modalities. For instance, the word *squealing* is highly evocative of a particular type of sound, and the link to this sound is more concrete due to the depictive qualities of this word. As already argued by Classen (1993, Chapter 3), this may prevent using such words outside of their core domains. The finding that iconicity is anti-correlated with crossmodal uses is in line with the general idea that iconicity is inimical to abstraction (Lupyan & Winter, 2018), which limits the reusability and extendibility of iconic word forms.

Finally, it should be emphasized that this chapter clearly demonstrated that to understand crossmodal language, one first needs a thorough understanding of sensory words more generally. In showing that the hierarchy of the senses is influenced by such factors as the composition of the lexicon, emotional valence, and iconicity, I have used findings from the previous chapters to shed light on crossmodal uses of sensory words. It seems that a lot of previous research in this literature has studied crossmodal language without considering the general properties of the sensory lexicon. The results presented in this chapter show that once one considers these properties, the idea that crossmodal language is governed by a monolithic hierarchy falls apart.

---

[1] Dimension words were based on manual classification, but similar results are obtained if the words belonging to the "shape & extent" cluster from Chapter 13 are excluded. It should further be noted that because most dimension words are classified as primarily "visual" by the native speakers in Lynott and Connell's (2009) study, including dimension words would only serve to strengthen the evidence against the hierarchy because more visual words would be sources (which goes together with a lower position on the hierarchy for this modality).

[2] These percentages obviously depend on what hierarchy is being assessed, which is not always specified in the literature on synesthetic metaphors.

[3] In fact, one might even want to argue that computing the proportion of hiearchy-consistent versus hierarchy-inconsistent cases is an analytical practice that is biased towards finding evidence for the hierarchy. Moreover, a gross proportion measure neglects the fact that given the linear ordering of the hierarchy, we actually *expect* some senses to be more frequent than others. In particular, touch should have a higher source-target ratio than taste, smell, sound, and sight, respectively. Obtaining source-target ratios that obey this order is actually more compelling

evidence for the hierarchy than mere counts of hierarchy-consistent versus hierarchy-inconsistent cases, which neglects the role of order.

[4] I chose to present separate models for valence and iconicity in the main text for ease of discussion. Moreover, since the data do not fully overlap, combining them leads to unnecessary exclusions. The online script includes several analyses that also control for additional factors, as well as analyses that simultaneously incorporate valence and iconicity. In the additional iconicity model (mixed logistic regression on categorical data), I controlled for auditory strength. In this case, the iconicity effect disappeared (+0.02, $SE = 0.13$, $p = 0.857$), which suggests that because of their strong correlation, effects in the simple iconicity-only model might actually driven by auditory strength, rather than iconicity per se. However, a separate subset analysis of auditory words only (dominant modality classification) shows a reliable effect of iconicity (–0.97, $SE = 0.002$, $p < 0.0001$). This suggests that even within the class of highly auditory words, differences between iconic and non-iconic words matter (e.g., *loud* versus *squealing*). Additionally, a separate analysis of cosines shows an iconicity effect even when the auditory strength of the adjective is controlled for (+0.02 higher cosines, $SE = 0.007$; $\chi^2(1) = 7.72$, $p = 0.005$).

There is a conflicting result for absolute context valence when gustatory and olfactory strength is controlled for in the categorical model (more valenced, *less* likely to be used in contexts); however, this effect reverted sign (in the predicted direction) in the cosine analysis that controlled for the same perceptual strength measures (–0.04, $SE = 0.01$; $\chi^2(1) = 12.6$, $p = 0.0004$). A simple linear model fitted on the average cosine per adjective and combining all factors (iconicity, valence, gustatory strength, olfactory strength, auditory strength) shows a reliable iconicity effect as well as a reliable context valence effect, both in the predicted direction.

Thus, overall the evidence is consistent with the predicted theoretical account, although it depends on model choice. Future research with extended iconicity ratings and modality ratings should see to replicate these results.

Chapter 18. Conclusion

18. Core themes

This chapter summarizes the book's core themes, including the five senses folk model (Chapter 18.1.2), the Embodied Lexicon Hypothesis (Chapter 18.1.2), metaphor (Chapter 18.1.3), ineffability and the composition of the English sensory vocabulary (Chapter 18.1.4), and methods (Chapter 18.1.5). I will then discuss possible applications (Chapter 18.2) and future directions (Chapter 18.3) for sensory linguistics.

18.1.1. The five senses folk model redux

In Chapter 2, I argued that the five senses folk model is a convenient falsehood, a temporary tool for making generalizations about the sensory vocabulary of English. The evidence presented in this book has shown the model both right and wrong. The model has proven its worth, for example, by showing us that vision is dominant in the English language (Chapters 12 & 15), that sound words are isolated from the rest of the sensory vocabulary (Chapters 12, 13, 14, & 17), and that taste and smell words are relatively more emotional (Chapter 16). These conclusions would not be possible without assuming the five sense folk model in some way or another.

On the other hand, I have also shown the limitations of assuming a five-fold distinction of the sensory world. Chapter 13 found that there are both more and less distinctions, depending on whether one assumes a micro- or a macro-perspective on the sensory vocabulary of English. When zooming in, one finds many more than just five categories. When zooming out, one finds fewer than five senses. Both of these perspectives are equally true.

The degree to which the five senses folk model can be misleading is perhaps most apparent for taste and smell. Although we distinguish these two senses in the five senses folk model, the data shows that taste and smell words are similar on almost any measure possible. This book showed that taste and smell words have similar overall perceptual strength ratings (Chapters 11–13); similar usage patterns in naturally occurring language (Chapter 14), including similar word frequencies (Chapter 15); and similar evaluative qualities (Chapter 16). Perhaps it is best to think of taste and smell words as forming one unified vocabulary. According to this view, there is no taste vocabulary that is clearly separated from the smell vocabulary. Instead, some words veer toward the taste pole and some toward the smell pole of an underlying taste–smell continuum. There are, however, some differences between words that are primarily gustatory and words that are primarily olfactory. In particular, on the taste end of the continuum, there is a higher preponderance of positive words (*delicious*, *tasty*, *sweet*). On the smell end of the continuum, there is a higher preponderance of negative words (*pungent*, *stinky*, *smelly*). That is, taste exhibits positive differentiation and smell exhibits negative differentiation within the English sensory vocabulary.

The five senses folk model also breaks down when considering the multisensoriality of perceptual words. The fact that sensory words are multisensory means that one has to be extremely careful about classifying sensory words in a hardcut fashion, which I only did sparingly in this book (namely, for isolated analyses in Chapter 12 and Chapter 17). The data presented in this book, particularly in Chapters 13 through 16, demonstrates that it may be misleading to shoehorn sensory words into discrete categories, such as when saying that *crunchy* is a touch word and only a touch word. Although such labels may provide useful heuristics in some cases, they generally neglect the underlying multisensoriality that characterizes perception as well as sensory language.

The results presented in this book highlight how it always must be kept in mind that when speakers are in the heat of a conversation, discrete categories such as "taste words" or "touch words" do not matter as much. In broad correspondence with a continuous view of language and cognition (e.g., Spivey, 2007), the results in Chapters 15 and 16 demonstrated how the continuous association of words to senses is more predictive of linguistic behavior than categorical classification systems. This is the case even when we use more very fine-grained categorical classifications such as the clusters computed in Chapter 13. I have argued repeatedly throughout this book, including in the discussion of crossmodal language (Chapter 7), that linguistic categories such as "taste words" or "touch words" only emerge when speakers, including linguists, reason *about* language in a metalinguistic fashion (compare Connell & Lynott, 2016). Actual language use may not be structured around these categories.

Whatever we conclude from these reflections on the five senses folk model, it should be clear that sensory linguistics cannot tacitly assume its validity. To the extent that the five senses model is employed in linguistic research, this needs to be openly addressed.

18.1.2. The Embodied Lexicon Hypothesis

What evidence presented in this book supports the Embodied Lexicon Hypothesis? Let us take stock of perception–language correspondences seen throughout this book.

First, vision is dominant in perception (Chapter 3) and in language. The dominance of visual language was revealed through multiple patterns, ranging from dominance in type frequencies (Chapter 12), through dominance in token frequencies (Chapter 15), all the way to crossmodal language, where vision was found to be a primary "target" of semantic extension (Chapter 17). Visual adjectives were furthermore found to be relatively high in semantic complexity (Chapter 15) but not emotionality (Chapter 16). The lack of specialization into emotionality is consistent with visual dominance: Whereas taste and smell are restricted to emotional language, vision is not. Taste and smell language is obligatorily evaluative (compare Levinson & Majid, 2014). On the other hand, visual language *can* express evaluation (*attractive*, *ugly*, *shiny*), but it does not have to since there are also many more neutral words to draw from. In Chapter 4, I discussed the fact that there is a multiplicity of factors which may lie behind visual dominance in language, such as cultural factors. However, given the strong evidence for visual dominance in perception, the available linguistic evidence is at least *consistent* with an embodied explanation and thus supports it indirectly.

A second correspondence between language and perception involves the fact that taste and smell are more emotional in perception, and the associated words are more emotional as well, compared to words from the other senses. Chapter 16 showed that the taste and smell adjectives from Lynott and Connell (2009) have overall higher emotional valence ratings. These results were obtained by looking at valence ratings in isolation (Warriner et al., 2013), as well as by looking at valence in contexts (Snefjella & Kuperman, 2016), which constituted a specific operationalization of the corpus-linguistic notion of semantic prosody. There is, by now, a wealth of converging evidence for the emotionality of taste and smell language (see Winter, 2016).

Third, I have shown that those modalities for which there is independent language-external evidence for intense crossmodal integration are also integrated in language (Chapters 13–14). In particular, taste and smell are highly associated in perception, which is mirrored by the fact that taste and smell are similarly inseparable in language. The same applies to the highly integrated modalities of sight and touch: There is abundant evidence for the neural and behavioral integration of these two modalities (Chapter 14), and this is mirrored in how sight and touch are associated with each other in the lexicon (Chapter 13) and in language use (Chapter 14).

Fourth, and perhaps most abstractly, I have discussed repeatedly that perception is intensely multisensory (e.g., Spence, 2011; Spence & Bayne, 2015), and so is language (Chapters 11–17). The multisensoriality of perception is reflected in the fact that the perceptual vocabulary of English is multisensory as well, although there are limits to this multisensoriality (Chapter 4). This is another correspondence between language and perception, albeit a fairly high-level one.

Together, these patterns can be seen as supporting the Embodied Lexicon Hypothesis. In addition, it should be kept in mind that there already is a lot of independent empirical evidence for the embodied processing of sensory words coming from psycholinguistic studies, reviewed in Chapter 5. The evidence presented here suggests that not only processing, but also structural and usage patterns follow embodied principles.

18.1.3. Metaphor

Crossmodal language use is a fascinating topic, rich with hypotheses that await to be tested (see Chapter 9). In Chapters 6 through 9, my discussion of crossmodal language was largely theoretical and argumentative, using our existing knowledge of synesthesia, metaphor, and the senses, to question the synesthetic and metaphorical nature of synesthetic metaphors. Whereas expressions such as *sweet smell* and *hot food* are generally seen as metaphors (involving taste-to-smell and temperature-to-taste mappings), Chapter 7 argued for seeing adjectives such as *sweet* and *hot* as highly supramodal descriptors that encompass multiple senses (Paradis & Eeg-Olofsson, 2013; Rakova, 2003).

My reconceptualization of synesthetic metaphors also stands against the five senses folk model, which assumes *separate* senses. Howes (2006) mentions how people generally think that "each sense has its own proper sphere" (p. 381). In contrast to this common belief, sensory words appear to involve highly overlapping spheres. Once we recognize this multisensoriality, crossmodal language appears less metaphorical. With Rakova (2003, p. 15), I concluded that "there is much more literalness in language than has traditionally been supposed" (see also Paradis & Eeg-Olofsson, 2013). A literal analysis of synesthetic metaphors has far-reaching conclusions for lexical semantics and conceptual metaphor theory (Gibbs, 1994; Kövecses, 2002; Lakoff, 1987; Lakoff & Johnson, 1980). In essence, a literal analysis compels us to see the continuity of the senses as reaching all the way down into the lexical representation of individual words. Such a view is, in fact, more embodied than assuming separate senses (Chapter 7).

Moreover, a literal analysis has an advantage in preventing us from positing an ever-growing number of metaphors. If too many disparate phenomena are subsumed under the same label "metaphor," any theory to account for these phenomena will become either unwieldy or hollow. Restricting the number of phenomena to which the notion of metaphor is applied helps to make metaphor theory theoretically and empirically tractable. In this book, I suggested that at least those linguistic expressions that are called "synesthetic metaphors" do not have to be accounted for by theories such as conceptual metaphor theory.

I furthermore discussed the fascinating proposal that a hierarchy of the senses governs how sensory words can be combined with each other (Chapters

8–9)—that is, some senses are more likely used to talk about other senses than the reverse. I have argued that we need to distinguish between the hierarchy as a descriptive phenomenon and the hierarchy as a governing principle. Although Chapter 17 found evidence that appears to be in line with the hierarchy, it only supports this as a descriptive generalization. A deeper look at the data suggests that there is no evidence for a monolithic hierarchy of the senses, such as Shen's directionality principle (Shen & Gil, 2007). In particular, I argued that it is important that one looks at which *specific* crossmodal combinations are overrepresented. When one does this, the pattern of results looks much more consistent with the Embodied Lexicon Hypothesis and the results presented in Chapter 14—that is, modalities that are perceptually associated with each other are also linguistically associated with each other. In addition, I showed that the hierarchy of the senses is inevitable given the composition of the word lists used in this study. The hierarchy of the senses may be grounded in the simple fact that some senses have more adjectives than others. The importance of what others call synesthetic metaphors is furthermore diminished when one acknowledges the fact that due to modality affinity (Chapter 14), there is a tendency for sensory words to be used in contexts that involve highly similar rather than dissimilar sensory modalities.

The idea of a monolithic hierarchy of the senses was further deconstructed by employing a multiexplanatory approach. Rather than assuming that the empirically observed asymmetries are caused by one principle, I argued that it is fruitful to look at how multiple factors are at play. Any discussion of the hierarchy of the senses needs to acknowledge that many different factors have been proposed, many of which have been not been tested yet (see Chapter 9). However, in Chapter 17, I showed new evidence consistent with the role of at least two linguistic factors, emotional valence and iconicity. In particular, corpus analyses revealed that highly evaluative adjectives were *more likely* to be used crossmodally. On the other hand, highly iconic adjectives were *less likely* to be used crossmodally. Future research needs to look at the influence of other linguistic factors. Ultimately, the goal is to have a multivariate model of crossmodal language use that combines the different factors and is able to assess their relative influence. However, given the evidence presented in Chapter 17 and the theoretical discussion in Chapter 9, it is clear that it is highly unlikely a one-size-fits-all principle lies behind the empirically observed asymmetries in crossmodal language use.

18.1.4. Ineffability and the composition of the sensory vocabulary

There are several results that speak to the high-level design characteristics of the English sensory vocabulary. I have shown that words occupy a sweet spot between complete unisensoriality on the one hand and complete multisensoriality on the other. Adjectives for perceptual sensations are more multisensory than most linguistic analyses acknowledge, but they also show clear signs of specialization (Chapter 12), including the fact that sensory words pair with words from their own modality in language use (Chapter 14).

Specialization leads to ineffability. Fainsilber and Ortony (1987) say that "language partitions the continuity of experience into discrete units comprised of words and phrases having a relatively narrow referential range" (p. 240). As was shown in Chapter 4, this entails that language is highly limited in its ability to convey fine perceptual detail. Moreover, the specialization of sensory words also means that language is incapable of expressing the subjective and multisensory nature of perceptual experience.

When it comes to the differential ineffability of the senses, I have presented new evidence for smell ineffability, as well as taste ineffability.

Ankerstein and Pereira (2013) showed that even when people describe food items, they list relatively few taste terms. The data presented in this book suggests that there are fewer unique taste and smell words (Chapter 12), and these words are used less frequently (Chapter 15). The fact that smell was infrequently a source but frequently a target of crossmodal expressions (Chapter 17) can be seen as further evidence for smell ineffability, given that "metaphor" is frequently seen as a device to express the inexpressible (Fainsilber & Ortony, 1987; Ortony, 1975). From this perspective, taste is slightly more effable, as it has dedicated taste words such as *sweet*, *sour*, and *bitter*, which are also frequently used to talk about words associated with the other senses. Thus, whereas both taste and smell lack lexical differentiation compared to sight, the ineffability of smell is more pronounced.

On the other hand, there were also weaker signs for the ineffability of sound. Throughout the book, sound was shown to have an isolated position in the sensory lexicon, with high exclusivity ratings (Chapter 12), anti-correlations with all other senses in naturally occurring language (Chapter 14), and a low propensity for semantic extension (Chapter 15), as well as a low propensity for being used to describe the other senses in crossmodal expressions (Chapter 17). Sound has relatively little dedicated vocabulary of its own (Chapter 12), in particular in the adjectival domain. Miller and Johnson-Laird (1976) also mention that "there are names for colors but no similar names for pitches" (p. 23). They similarly say that "for timbre there is a host of ill-defined terms" (p. 24), many of which are metaphorical. Miller and Johnson-Laird conclude that "it is ironic that people use vocal sounds to name everything else yet have such a limited vocabulary for sounds themselves" (p. 25).

In line with these indicators of ineffability, sound has been found to be a frequent target of crossmodal expressions in Chapter 17. The fact that the expression of sound needs support from "metaphor" has been noted by many researchers. Engstrom (1946, pp. 10–11) notes in passing that there are particularly many metaphors for descriptions of voices. Pérez-Sobrino and Julich (2014) find that music descriptions have more metaphorical content than other types of discourse.

This book has also shown something previously only demonstrated for signed languages (Emmorey, 2014; Meir, 2010)—namely, that iconicity restricts semantic extension. This is one additional factor that explains the low propensity of sound concepts being used to talk about the other senses. The high iconicity of sound words (Chapter 15) may restrict sound language even more (Chapter 17). The finding that iconicity restricts semantic extendibility is also an important result for iconicity research because it may be that this restrictive nature of iconicity is one reason for why languages are not more iconic than they could be (Lupyan & Winter, 2018). Moreover, this finding is important for linguistics more generally because it shows that the nature of the form–meaning mapping of a word has downstream effects in the linguistic system, such as in which linguistic contexts a word occurs.

18.1.5. Methods

Finally, a theme that ran across the entire book was the use of rigorous statistical methods, coupled with highly constrained operational definitions. Chapter 11 defended the validity of introspective data, as long as introspective judgments are collected from a large number of unbiased participants. The resulting semantic norms can be incorporated into a corpus-based work flow to counteract the fact that the ratings were collected in an isolated and decontextualized fashion.

All empirical chapters demonstrated how useful rating studies are for sensory linguistics. This usefulness extends to the language sciences writ large, in particular to the quantitative study of word meaning. Throughout the book, I have used the modality norms by Lynott and Connell (2009), first introduced in Chapter 10. It was shown that given the right analyses, there is a wealth of theoretically important findings that can be generated with such a simple dataset. Chapters 12 and 13 showed that the ratings can be studied in their own right. Chapter 14 and Chapter 17 showed that the ratings can be studied together with corpus data. Chapters 15, 16, and 17 showed that the modality norms can be combined with a wealth of linguistic resources (including corpus frequencies and dictionary meaning counts) and other norm datasets (valence norms and iconicity norms). This illustrates the utility of a norm-based linguistics. In the future, the availability of norms from numerous large-scale rating studies will prove to be an invaluable addition to the linguistic toolkit.

The methods used throughout this book furthermore allowed the quantification of two core concepts of corpus linguistics: Chapter 14 provided quantitative evidence for the notion of semantic preference, and Chapter 16 achieved the same for the related notion of semantic prosody. Interestingly, in *both* cases, the evidence showed that ratings on isolated words are highly predictive of linguistic behavior in context. With respect to semantic preference, it was found that the crossmodal correlations in the rating data (Chapter 13) correspond to crossmodal correlations observed in corpora (Chapter 14). With respect to semantic prosody, I demonstrated a close fit between decontextualized valence ratings and the valence of contexts in corpus data (Chapter 16). Corpus linguistics has an aversion to introspective judgments on isolated words, often for the right reasons. However, the results from this book clearly show that in some cases, considerable leverage is gained from looking at words without context.

There were several methodological themes, first introduced in Chapter 10, that crosscut all of the analyses presented here. These include the issue of reproducibility and how important it is to utilize explicit criteria, as well as how important it is to publish one's data and analysis code. As many analytical decisions as possible need to be laid open to be transparent to other researchers, and data and methods need to be shared to achieve cumulative progress in the language sciences. In line with the theme of reproducibility, data and code are publically available for all analyses presented in this book. Moreover, as argued in Chapter 10, the use of standardized norm data furthers the reproducibility of semantic analyses, as different researchers will come to the same conclusion when given the same dataset.

Another methodological theme that characterized multiple chapters can be related to the phenomenon of "base rate neglect" (see Pennycook & Thompson, 2017), whereby people (including researchers) ignore or undervalue the a priori probability of a particular phenomenon. I would argue that such base rate neglect characterized the previous literature on synesthetic metaphors in two ways. First, the phenomenon of crossmodal language was not compared to the base rate of unimodal language, which I explored under the concept of "modality affinity" in Chapter 14, as well as by quantifying the number of within-modality uses of sensory words in Chapter 17. When this base rate is considered, one has to acknowledge that genuinely crossmodal language is not all that frequent compared to the within-modality uses of sensory words. Second, when researchers investigated the hierarchy of the senses, they did not consider the base rate of sensory words for particular sensory modalities as much. In Chapter 17, I showed that the composition of the sensory lexicon affects the types of hierarchical patterns we see in corpus data. More generally, I argued that any analysis of crossmodal language use (such as is done in the literature on synesthetic metaphors) needs to be based on a thorough understanding of the sensory lexicon.

The idea of a "base rate" was also an analysis strategy in other chapters. In Chapter 12, I reinvestigated Lynott and Connell's (2009) claim that a modality exclusivity of 46% means that sensory words are multisensory. By reshuffling the rating data (permutation-based approach), I was able to create a base line against which to compare this figure. This analysis showed that despite the evidence for multisensoriality, sensory words exhibit statistically reliable specialization into particular sensory modalities (i.e., a drive towards unisensoriality). In Chapter 14, I created a base line of crossmodal language use by combining all Lynott and Connell (2009) adjectives with all Lynott and Connell (2013) nouns. These hypothetical adjective–noun pairs were then compared with corpus data to show that sensory words tend to stick to their own kind, or to highly related sensory modalities (modality affinity). Altogether, these analyses suggest that quantitative research on perceptual language needs to think more often and more deeply about whether an observed phenomenon is actually unusual. It is then important to create statistical base lines to act as appropriate points of comparison.

A final methodological theme that characterized all analyses was that imposing constraints on oneself is often a good thing. Chapter 10 argued for endorsing abstraction and simplification to achieve higher levels of generalization and falsification. In line with Healy's (2017) "fuck nuance" credo, the analyses showed that considerable leverage is to be gained, for example, by assuming the obviously oversimplified model of the five senses, or by restricting the complex notion of emotional meaning to the positive/negative dimension alone. As long as simplification is done carefully and knowingly, and as long as it is addressed openly, it can further progress the language sciences.

18.2. Applications

This book focused on fundamental problems of sensory language, including the general question of how perceptual content is encoded in language. Given the central function of the senses in our daily lives and in society, the issues raised here naturally have far-reaching implications for applications.

Perhaps the most obvious domain of application is in advertising and marketing (see also Fenko et al., 2010). Several researchers have argued for the importance of what is called “sensory marketing” (Hultén, 2015; Hultén, Broweus, & van Dijk, 2009; Lindstrom, 2010), an approach to marketing that focuses on engaging all the senses. In particular, sensory marketing pushes against the fact that to this day, most advertising focuses almost exclusively on sight and sound alone. Elder and Krishna (2009) showed that ads describing products with multisensory language are more effective than ads that focus on only one sense. This research substantially benefits from a deep theoretical understanding of sensory language. This research also benefits from the methods discussed here. For instance, Lynott and Connell’s (2009) dataset could be used to select sensory adjectives for advertisements in a more systematic fashion, and to quantify the degree of multisensoriality that is used in a specific advertisement. More generally, knowing about the semiotic toolkit (Chapter 3) and the limits of language (Chapter 4) is key to advertisers who frequently need to find effective ways of communicating the sensory qualities of products.

The book’s findings also have applications in food science, where researchers work together with practitioners to develop sensory lexicons to standardize the description of such domains as wine, coffee, or kimchi. These standards are used (among other things) in food testing, where it is important to have a shared vocabulary between experts. Knowing about sensory language (and its limitations) is key to developing effective sensory vocabularies, as is investigating how expert vocabularies differ from lay vocabularies (e.g., Diederich, 2015).

Perceptual psychology research also benefits from a firm understanding of the properties of the sensory lexicon. In many perceptual psychology experiments, responses are delivered verbally. Evidence for the role of language in perceptual tasks has been attained by Huisman and Majid (2018), who showed that odors with high-frequency labels were named correctly more often. This suggests that when perception is studied using linguistic tasks, the researcher has to pay attention to the linguistic properties of the words involved, such as the many properties discussed throughout this book. In fact, the perceptual researcher must consider the fact that had the perceptual task been done in other languages, different results may have been obtained. Evidence for this comes from the fact that the deficiency of smell naming is connected to the relative lack of smell terms in English, at least when compared to the vocabularies of other languages with larger smell vocabularies, such as the Jahai from Malaysia (Majid & Burenhult, 2014).

Finally, the research presented in this book may have relevance for research into visual impairment and other sensory deficits. For example, many researchers are interested in audio descriptions for the visually impaired (e.g., Peli, Fine, & Labianca, 1996; Schmeidler & Kirchner, 2001; Szarkowska, 2011). To develop effective audio descriptions, it helps to understand not only how sensory language works, but also how the sensory language of sighted individuals differs from the sensory language of individuals with visual impairments (see Landau & Gleitman, 1985; Shephard & Cooper, 2001).

Thus, although the book’s goals were primarily theoretical, a thorough understanding of sensory language is very relevant to a whole range of applied domains.

18.3. Future directions

Although Part II of this book has reported a large number of results on perceptual language, there are many things that we currently do not know. Sensory linguistics is rich with hypotheses waiting to be tested and issues waiting to be explored. The following list presents range of potential topics for future research that is intended to pique interest.

- Dąbrowska (2016a) lists the neglect of individual differences as one of the “sins” of cognitive linguistics. For sensory language, there is a whole world of individual differences to explore. Do people who are better tasters or better smellers also have different taste and smell vocabularies? What about experts such as wine tasters and coffee tasters? (see Croijmans & Majid, 2016)

- Is the sensory language of blind individuals different from the sensory language of sighted individuals (see, e.g., Landau & Gleitman, 1985)? Do blind individuals, for example, use touch-related adjectives in a different fashion?

- What about deaf individuals? What semiotic strategies are used to talk about such perceptual domains as sound, smell, and color? Are there noteworthy differences in the sensory vocabularies of different signed languages?

- Do synesthetes use sensory language differently? Do they use different crossmodal expressions? At present, it is impossible to make inferences about a person's perceptual system from language use alone (see Cytowic & Eagleman, 2009, p. 188), but at some point such inferences may be possible.

- How does the claim that synesthetic metaphors are best analyzed as literal expressions (Chapter 8) pare against classic approaches within lexical semantics, such as tests of ambiguity and vagueness? (Geeraerts, 1993; Zhang, 1998; Zwicky & Sadock, 1975)

- Do strong and weak synesthetic metaphors (*cold anger* versus *cold sound*) share an underlying mechanism, as hinted at by the results of Chapter 14 and Chapter 17?

- There are many different explanations for the hierarchy of the senses (Chapter 9). What explanatory factors matter the most in explaining asymmetries between crossmodal uses of sensory adjectives? How do the different explanatory factors (frequency, iconicity, gradability, evaluation, etc.) interact?

- How do the relations between sensory words change between different types of discourse? For example, does the hierarchy of the senses differ between such text types as perfume descriptions, wine tasting notes, and music reviews?

- Do speakers of languages with more elaborate smell vocabularies (such as the Austroasiatic languages Maniq and Jahai; Majid & Burenhult, 2014, Majid et al., 2018; Wnuk & Majid, 2014; or the Austronesian language

Amis; Lee, 2015) use crossmodal expressions differently? Investigating this is particularly important because the evidence for the hierarchy of the senses has so far come almost exclusively from languages spoken in large industrialized societies.

- How are sensory words used together with other semiotic tools in situated interactions, such as when buying cheese in an artisan cheese shop or when buying perfume?

- How do the results obtained here for sensory adjectives compare to other parts of speech, such as nouns and verbs?

- What happens when sensory words are translated between languages? How much perceptual information is carried over in translation? Can modality norms be translated?

- How does sensory language interact with gesture?

- The analyses presented here are exclusively synchronic and need to be supplemented with diachronic analyses. How does sensory language change over time?

- How do technologies, societal practices, and belief systems shape sensory language?

These are but some of the many questions that fall within the domain of sensory linguistics. As is clear from the list, there are many questions that await to be answered. Sensory linguistics is a thriving field with much left to be explored.

18.4. Conclusions

Clearly, “we are not just minds floating in the air” (Rakova, 2003, p. 18). Humans are bodily beings who are connected to the world around them via their senses. Far from the frequently held view that language is an arbitrary symbol system that mostly obeys abstract formal principles, this book has demonstrated that language is deeply infused with sensory information, and knowing about our sensory world helps us understand the linguistic world we live in. Moreover, language provides a window into the senses. Using sensory linguistics, we can study what Marks (1978) has called “the fabric of mental tapestry richly woven in form and color, sound, taste, touch, and scent” (p. 255). Using the tools outlined here, we can study some of the most fundamental aspects of being human: namely, how we sense the world around us, and how we can communicate our sensory worlds to others.

References

Abraham, W. (1987). Synästhesie als Metapher. *Folia Linguistica*, *21*, 155-190.

Agapakis, C. M., & Tolaas, S. (2012). Smelling in multiple dimensions. *Current Opinion in Chemical Biology*, *16*, 569-575.

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *1*, 298-348.

Akpinar, E. & Berger, J. (2015). Drivers of cultural success: The case of sensory metaphors. *Journal of Personality and Social Psychology*, *109*, 20-34.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14, 257-262.

Alivisatos, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital cortex. *Cerebral Cortex*, *12*, 1202-1212.

Allan, K., & Burridge, K. (2006). *Forbidden words: Taboo and the censoring of language*. Cambridge, UK: Cambridge University Press.

Anaki, D., & Henik, A. (2017). Bidirectionality in synesthesia and metaphor. *Poetics Today*, *38*, 141-161.

Amassian, V. E., Cracco, R. Q., Maccabee, P. J., Cracco, J. B., Rudell, A., & Eberle, L. (1989). Suppression of visual perception by magnetic coil stimulation of human occipital cortex. *Electroencephalography and Clinical Neurophysiology*, *74*, 458-462.

Amsel, B. D., Urbach, T. P., & Kutas, M. (2012). Perceptual and motor attribute ratings for 559 object concepts. *Behavior Research Methods*, *44*, 1028-1041.

Ankerstein, C. A., Pereira, G. M. (2013). Talking about taste: Starved for words. In C. Gerhardt, M. Frobenius, & S. Ley (Eds.), *Culinary linguistics: The chef's special* (pp. 305-315). Amsterdam: John Benjamins.

de Araujo, I. E., Rolls, E. T., Kringelbach, M. L., McGlone, F., & Phillips, N. (2003). Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. *European Journal of Neuroscience*, *18*, 2059-2068.

Arshamian, A., & Larsson, M. (2014). Same same but different: the case of olfactory imagery. *Frontiers in Psychology*, *5*, 34.

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross - linguistic comparison of English and Swedish speakers. *Cognitive Science*, *37*, 286-309.

Auvray, M., & Spence, C. (2008). The multisensory perception of flavor. *Consciousness and Cognition*, *17*, 1016-1031.

Baayen, R. H., & del Prado Martín, F. M. (2005). Semantic density and past-tense formation in three Germanic languages. *Language*, *81*, 666-698.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the 7th Conference on Language Resources and Evaluation*, *10*, 2200-2204.

Backhouse, A.E. (1994). *The lexical field of taste: a semantic study of Japanese taste terms*. Cambridge, UK: Cambridge University Press.

Bagli, M. (2016). "Shaking off so good a wife and so sweet a lady": Shakespeare's use of taste words. *Journal of Literary Semantics*, *45*, 141-159.

Bagli, M. (2017). Tastes we've lived by: Taste metaphors in English. *Textus*, *30*, 33-48.

Baker, S. J. (1950). The pattern of language. *The Journal of General Psychology*, *42*, 25-66.

Barrós-Loscertales, A., González, J., Pulvermüller, F., Ventura-Campos, N., Bustamante, J. C., Costumero, V., Parcet, M. A., & Ávila, C. (2011). Reading salt activates gustatory brain regions: fMRI evidence for semantic grounding in a novel sensory modality. *Cerebral Cortex*, 22, 2554-2563.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Barten, S. S. (1998). Speaking of music: The use of motor-affective metaphors in music instruction. *Journal of Aesthetic Education*, *32*, 89-97.

Bartley, S. H. (1953). The perception of size or distance based on tactile and kinesthetic data. *The Journal of Psychology*, *36*, 401-408.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015a). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-9.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015b). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, *67*, 1-48.

Baumann, O., & Greenlee, M. W. (2007). Neural correlates of coherent audiovisual motion perception. *Cerebral Cortex*, *17*, 1433-1443.

Baumgärtner, K. (1969). Synästhesie und das Problem sprachlicher Universalien. *Zeitschrift für deutsche Sprache*, *25*, 1-20.

Basbaum, A. I., Bautista, D. M., Scherrer, G., & Julius, D. (2009). Cellular and molecular mechanisms of pain. *Cell*, *139*, 267-284.

Bednarek, M. A. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory*, *4*, 119-139.

Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning*. New York: Basic Books.

Berglund, B., Berglund, U., Engen, T., & Ekman, G. (1973). Multidimensional analysis of twenty-one odors. *Scandinavian Journal of Psychology*, *14*, 131-137.

Berlin, B. (2006). The first congress of ethnozoological nomenclature. *Journal of the Royal Anthropological Institute*, *12*, S23-S44.

Berlin, B., & O'Neill, J. P. (1981). The pervasiveness of onomatopoeia in Aguaruna and Huambisa bird names. *Journal of Ethnobiology*, *1*, 238-261.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Bhushan, N., Rao, A. R., & Lohse, G. L. (1997). The texture lexicon: Understanding the categorization of visual texture terms and their relationship to texture images. *Cognitive Science*, *21*, 219-246.

Birch, J. (2012). Worldwide prevalence of red-green color deficiency. *Journal of the Optical Society of America A*, *29*, 313-320.

Blake, R., Sobel, K. V., & James, T. W. (2004). Neural synergy between kinetic vision and touch. *Psychological Science*, *15*, 397-402.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*, 10818-10823.

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science*, *13*, 185-188.

Bottini, R. & Casasanto, D. (2010). Implicit spatial length modulates time estimates, but not vice versa. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 1348-1353). Austin, TX: Cognitive Science Society.

Box, G. E. (1979). Robustness in the strategy of scientific model building. *Robustness in Statistics*, *1*, 201-236.

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition*, *126*, 165-172.

Brower, D. (1947). The experimental study of imagery: II. The relative predominance of various imagery modalities. *The Journal of General Psychology*, *37*, 199-200.

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, *49*, 454-462.

Brunsdon, C., & Chen, H. (2014). GISTools: Some further GIS capabilities for R. R package version 0.7-4. https://CRAN.R-project.org/package=GISTools

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977-990.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*, 991-997.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904-911.

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *12*, 1831-1840.

Buck, C. D. (1949). *A dictionary of selected synonyms in the principal Indo-European languages: A contribution to the history of ideas*. Chicago: University of Chicago Press.

Bylund, E., Athanasopoulos, P., & Oostendorp, M. (2013). Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics*, *51*, 929-955.

Caballero, R. (2007). Manner-of-motion verbs in wine description. *Journal of Pragmatics*, 39, 2095-2114.

Caballero, R., & Ibarretxe-Antuñano, I. (2014). Ways of perceiving, moving, and thinking: Revindicating culture in conceptual metaphor research. *Cognitive Semiotics*, *V*, 268-290.

Caballero, R., & Paradis, C. (2015). Making sense of sensory perceptions across languages and cultures. *Functions of Language*, *22*, 1-19.

Cabanac, M. (1971). Physiological role of pleasure. *Science*, *173*, 1103-1107.

Cabanac, M., Pruvost, M., & Fantino, M. (1973). Alliesthesie negative pour des stimulus sucres apres diverses ingestions de glucose. *Physiology & Behavior*, *11*, 345-348.

Caclin, A., Soto-Faraco, S., Kingstone, A., & Spence, C. (2002). Tactile "capture" of audition. *Attention, Perception, & Psychophysics*, *64*, 616-630.

Cain, W. S. (1979). To know with the nose: keys to odor identification. *Science*, *203*, 467-470.

Campbell, J. D., & Katz, A. N. (2006). On reversing the topics and vehicles of metaphor. *Metaphor and Symbol*, *21*, 1-22.

Canobbio, A. T. (2004). *Trattamento delle sensazioni nella poesia di Guido Gozzano*. Genova: University of Genova MA thesis.

Caplan, D. (1973). A note on the abstract readings of verbs of perception. *Cognition*, *2*, 269-277.

Carlson, N. R. (2010). *Physiology of behavior* (10th Ed.). Boston: Allyn & Bacon.

Casagrande, V. A. (1994). A third parallel visual pathway to primate area V1. *Trends in Neurosciences*, *17*, 305-310.

Casati, R., Dokic, J., & Le Corre, F. (2015). Distinguishing the commonsense senses. In D. Stokes, M. Matthen & S. Biggs (Eds.), *Perception and its modalities* (pp. 462-479). Oxford: Oxford University Press.

Casasanto, D. (2014). Experiential origins of mental metaphors: Language, culture, and the body. In M. Landau, M.D. Robinson, & B. Meier (Eds.), *The power of metaphor: Examining its influence on social life* (pp. 249-268). Washington, DC: American Psychological Association Books.

Casasanto, D. (2017). The hierarchical structure of mental metaphors. In B. Hampe (Ed.), *Metaphor: Embodied cognition and discourse* (pp. 46-61). Cambridge, UK: Cambridge University Press.

Casasanto, D., & Jasmin, K. (2010). Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *PLOS ONE*, *5*, e11805.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*, 579-593.

Casasanto, D., & Chrysikou, E.G. (2011). When left is "right": Motor fluency shapes abstract concepts. *Psychological Science*, *22*, 419-422.

Cazeaux, C. (2002). Metaphor and the categorization of the senses. *Metaphor and Symbol*, *17*, 3-26.

Chambers, E., Lee, J., Chun, S., & Miller, A. E. (2012). Development of a lexicon for commercially available cabbage (baechu) kimchi. *Journal of Sensory Studies*, *27*, 511-518.

Crisinel, A. S., Cosser, S., King, S., Jones, R., Petrie, J., & Spence, C. (2012). A bittersweet symphony: Systematically modulating the taste of food by changing the sonic properties of the soundtrack playing in the background. *Food Quality and Preference*, *24*, 201-204.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*, 489-509.

Chu, S., & Downes, J. J. (2000). Long live Proust: The odour-cued autobiographical memory bump. *Cognition*, *75*, B41-B50.

Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, *122*, 211-226.

Citron, F. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience*, *26*, 2585-2595.

Citron, F. M., Güsten, J., Michaelis, N., & Goldberg, A. E. (2016). Conventional metaphors in longer passages evoke affective brain response. *NeuroImage*, *139*, 218-230.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 243-268). Hillsdale, NJ: Erlbaum.

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, *66*, 764-805.

Classen, C. (1993). *Worlds of sense: Exploring the senses in history and across cultures.* London: Routledge.

Classen, C. (1997). Foundations for an anthropology of the senses. *International Social Science Journal*, *49*, 401-412.

Classen, C., Howes, D., & Synnott, A. (1994). *Aroma: The cultural history of smell*. London: Routledge.

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., & Anderson, A. K., Rosen, B. R., & Belliveau, J. W. (1996). Changes in the cortical activity during mental rotation, a mapping study using functional MRI. *Brain*, *119*, 89-100.

Connell, L. (2007). Representing object colour in language comprehension. *Cognition*, *102*, 476-485.

Connell, L., & Lynott, D. (2009). Is a bear white in the woods? Parallel representation of implied object color during language comprehension. *Psychonomic Bulletin & Review*, *16*, 573-577.

Connell, L., & Lynott, D. (2010). Look but don't touch: Tactile disadvantage in processing modality-specific words. *Cognition*, *115*, 1-9.

Connell, L., & Lynott, D. (2011). Modality switching costs emerge in concept creation as well as retrieval. *Cognitive Science*, *35*, 763-778.

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, *125*, 452-465.

Connell, L., & Lynott, D. (2014). I see/hear what you mean: semantic activation in visual word recognition depends on perceptual attention. *Journal of Experimental Psychology: General*, *143*, 527-533.

Connell, L., & Lynott, D. (2016). Do we know what we're simulating? Information loss on transferring unconscious perceptual simulation to conscious imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1218-1232.

Connell, L. M., Lynott, D. J., & Banks, B. (2018). Interoception: the forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *373*, 20170143.

Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, *28*, 117-139.

Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks: Sage Publishing.

Craig, A. D. (2003). A new view of pain as a homeostatic emotion. *Trends in Neurosciences*, *26*, 303-307.

Crisinel, A. S., Jones, S., & Spence, C. (2012). 'The sweet taste of maluma': Crossmodal associations between tastes and words. *Chemosensory Perception*, *5*, 266-273.

Croijmans, I., & Majid, A. (2016). Not all flavor expertise is equal: The language of wine and coffee experts. *PLOS ONE*, *11*, e0155845.

Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge, UK: Cambridge University Press.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.

Curcio, C. A., Sloan Jr., K. R., Packer, O., Hendrickson, A. E., & Kalina, R. E. (1987). Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science*, *236*, 579-583.

Cuskley, C. (2013). Mappings between linguistic sound and motion. *Public Journal of Semiotics*, *5*, 39-62.

Cuskley, C., & Kirby, S. (2013). Synaesthesia, cross-modality and language evolution. In Simner, J. & Hubbard E.M. (Eds), *Oxford handbook of synaesthesia* (pp. 869-907). Oxford: Oxford University Press.

Cytowic, R. E., & Eagleman, D. (2009). *Wednesday is indigo blue: Discovering the brain of synesthesia*. Cambridge, MA: MIT Press.

Dąbrowska, E. (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review*, *27*, 1-23.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*, 219-253.

Dąbrowska, E. (2015). Individual differences in grammatical knowledge. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 649-667). Berlin: de Gruyter.

Dąbrowska, E. (2016a). Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics*, *27*, 479-491.

Dąbrowska, E. (2016b). Looking into introspection. In G. Drożdż (Ed.), *Studies in lexicogrammar: Theory and applications* (pp. 55-74). Amsterdam: John Benjamins.

Dacremont, C. (1995). Spectral composition of eating sounds generated by crispy, crunchy and crackly foods. *Journal of Texture Studies*, *26*, 27-43.

Dalton, P., Doolittle, N., Nagata, H., & Breslin, P. A. S. (2000). The merging of the senses: integration of subthreshold taste and smell. *Nature Neuroscience*, *3*, 431-432.

Dam-Jensen, H., & Zethsen, K. K. (2007). Pragmatic patterns and the lexical system—A reassessment of evaluation in language. *Journal of Pragmatics*, *39*, 1608-1623.

Dancygier, B., & Sweetser, E. (2014). *Figurative Language*. Cambridge, UK: Cambridge University Press.

van Dantzig, S., Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A sharp image or a sharp knife: Norms for the modality-exclusivity of 774 concept-property items. *Behavior Research Methods*, *43*, 145-154.

van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science*, *32*, 579-590.

Davies, M. (2008) The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at http://corpus.byu.edu/coca/

David, S. (1997). Représentations sensorielles et marques de la personne: contrastes entre olfaction et audition. In D. Dubois (Ed.), *Catégorisation et cognition: de la perception au discours* (pp. 209-242). Paris: Editions Kimé.

Davis, R. (1961). The fitness of names to drawings: A cross-cultural study in Tanganyika. *British Journal of Psychology*, *52*, 259-268.

Day, S. (1996). Synaesthesia and synaesthetic metaphors. *Psyche*, *2*, 1-16.

de Houwer, J., & Randell, T. (2004). Robust affective priming effects in a conditional pronunciation task: evidence for the semantic representation of evaluative information. *Cognition & Emotion*, *18*, 251-264.

Delwiche, J. F., & Heffelfinger, A. L. (2005). Cross-modal additivity of taste and smell. *Journal of Sensory Studies*, *20*, 512-525.

Dennett, D. C. (2013). *Intuition pumps and other tools for thinking*. London: Penguin Boos.

Deroy, O., & Spence, C. (2013). Why we are not all synesthetes (not even weakly so). *Psychonomic Bulletin & Review*, *20*, 643-664.

de Wijk, R. A., & Cain, W. S. (1994). Odor quality: discrimination versus free and cued identification. *Perception & Psychophysics*, *56*, 12-18.

Diederich, C. (2015). *Sensory adjectives in the discourse of food: A frame-semantic approach to language and perception*. Amsterdam: John Benjamins.

Dijksterhuis, G., Luyten, H., de Wijk, R., & Mojet, J. (2007). A new sensory vocabulary for crisp and crunchy dry model foods. *Food Quality and Preference*, i, 37-50.

Diffloth, G. (1994). i: big, a: small. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 107-114). Cambridge, UK: Cambridge University Press.

Dilts, P., & Newman, J. (2006). A note on quantifying 'good' and 'bad' prosodies. *Corpus Linguistics and Linguistic Theory*, *2*, 233-242.

Dingemanse, M. (2009). The selective advantage of body-part terms. *Journal of Pragmatics*, *41*, 2130-2136.

Dingemanse, M. (2011). *The meaning and use of ideophones in Siwu*. PhD dissertation. Radboud University, Nijmegen.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, *6*, 654-672.

Dingemanse, M. (2013). Ideophones and gesture in everyday speech. *Gesture*, *13*, 143-165.

Dingemanse, M. (2018). Redrawing the margins of language: Lessons from research on ideophones. *Glossa: A Journal of General Linguistics, 3*, 4.

Dingemanse, M., & Akita, K. (2017). An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics*, *53*, 501-532.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*, 603-615.

Dingemanse, M., & Majid, A. (2012). The semantic structure of sensory vocabulary in an African language. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 300-305). Austin, TX: Cognitive Science Society.

Dirven, R. (1985). Metaphor as a basic means for extending the lexicon. In W. Paprotté & R. Dirven (Eds.), *The ubiquity of metaphor: Metaphor in language and thought* (pp. 85-119). Amsterdam: John Benjamins.

Djordjevic, J., Lundstrom, J. N., Clement, F., Boyle, J. A., Pouliot, S., & Jones-Gotman, M. (2008). A rose by any other name: would it smell as sweet?. *Journal of Neurophysiology*, *99*, 386-393.

Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological Science*, *24*, 613-621.

Dominy, N. J., Lucas, P. W., Osorio, D., & Yamashita, N. (2001). The sensory ecology of primate food perception. *Evolutionary Anthropology*, *10*, 171-186.

Drury, H. A., Van Essen, D. C., Anderson, C. H., Lee, C. W., Coogan, T. A., & Lewis, J. W. (1996). Computerized mappings of the cerebral cortex: a multiresolution flattening method and a surface-based coordinate system. *Journal of Cognitive Neuroscience*, *8*, 1-28.

Dubois, D. (2000). Categories as acts of meaning: The case of categories in olfaction and audition. *Cognitive Science Quarterly*, *1*, 35-68.

Dunn, J. (2015). Modeling abstractness and metaphoricity. *Metaphor and Symbol*, *30*, 259-289.

Edwards, R., & Clevenger, T. (1990). The effects of schematic and affective processes on metaphorical invention. *Journal of Psycholinguistic Research*, *19*, 91-102.

Eklundh, L., & Singh, A. (1993). A comparative analysis of standardised and unstandardised principal components analysis in remote sensing. *International Journal of Remote Sensing*, *14*, 1359-1370.

Elder, R. S., & Krishna, A. (2009). The effects of advertising copy on sensory thoughts and perceived taste. *Journal of Consumer Research*, *36*, 748-756.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*, 301-306.

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*, 20130301.

Enfield, N. J. (2008). *The anatomy of meaning*. Cambridge, UK: Cambridge University Press.

Engen, T., & Ross, B. M. (1973). Long-term memory of odors with and without verbal descriptions. *Journal of Experimental Psychology*, *100*, 221-227.

Engstrom, A. G. (1946). In defence of synaesthesia in literature. *Philological Quarterly*, *25*, 1-19.

Erzsébet, P. D. (1974). Synaesthesia and poetry. *Poetics*, *3*, 23-44.

Essegbey, J. (2013). Touch Ideophones in Nyagbo. In O. O. Orie, & K. W. Sanders (Eds.), *Selected Proceedings of the 43rd Annual Conference on African Linguistics* (pp. 235-243). Somerville, MA: Cascadilla Proceedings Project.

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of 5th Conference on Language Resources and Evaluation* (Vol. 6, pp. 417-422).

Etzi, R., Spence, C., Zampini, M., & Gallace, A. (2016). When sandpaper is 'kiki' and satin is 'bouba': An exploration of the associations between words, emotional states, and the tactile attributes of everyday materials. *Multisensory Research*, *29*, 133-155.

Evans, N., & Wilkins, D. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*, *76*, 546-592.

Evans, V. (2007). *A glossary of cognitive linguistics*. Edinburgh: Edinburgh University Press.

Evans, V., & Green, M. (2006). *Cognitive linguistics: An introduction*. Mahwah: Lawrence Erlbaum Associates Publishers.

Evans, V., Bergen, B., & Zinken, J. (2007). The cognitive linguistics enterprise: An overview. In V. Evans, B. Bergen & J. Zinken (Eds.), *The cognitive linguistics reader* (pp. 2-30). London: Equinox.

Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology*, *46*, 621-646.

Fainsilber, L., & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbol*, *2*, 239-250.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229-238.

Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some wh-constraints in German. *Lingua*, *115*, 1525-1550.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Fenko, A., Otten, J. J., & Schifferstein, H. N. (2010). Describing product experience in different languages: The role of sensory modalities. *Journal of Pragmatics*, *42*, 3314-3327.

Fiedler, K. (2017). What constitutes strong psychological science? The (neglected) role of diagnosticity and a priori theorizing. *Perspectives on Psychological Science*, *12*, 46-61.

Fitch, T. (1994). Vocal tract length perception and the evolution of language. B.A. Thesis, Brown University.

Firth, J. R. (1957). *Papers in linguistics 1934–1951*. London: Oxford University Press.

Fischer, A. (1999). What, if anything, is phonological iconicity? In O. Fischer & M. Nänny (Eds.), *Form miming meaning: Iconicity in language and literature* (pp. 123-134). Amsterdam: John Benjamins.

Fischer, S. (1922). Über das Entstehen und Verstehen von Namen. *Archiv für die gesamte Psychologie*, *42*, 335-368.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61*, 825-850.

Floyd, S., San Roque, L., & Majid, A. (2018). Smell is coded in grammar and frequent in discourse: Cha'palaa olfactory language in cross - linguistic perspective. *Journal of Linguistic Anthropology*, *28*, 175-196.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Fontana, F. (2013). Association of haptic trajectories with takete and maluma. In I. Oakley, & S. Brewster (Eds.), *Haptic and audio interaction Design* (pp. 60-68). Berlin: Springer.

Forceville, C. (2006). Non-verbal and multimodal metaphor in a cognitivist framework: agendas for research. In G. Kristiansen, M. Achard, R. Dirven, F. Ruiz de Mendoza Ibañez (Eds.), *Cognitive linguistics: Current applications and future perspectives* (pp. 379–402). Berlin: de Gruyter.

Forceville, C. (2008). Metaphor in pictures and multimodal representations. In R. W. Gibbs (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 462–482). Cambridge, UK: Cambridge University Press.

Forceville, C. J., & Urios-Aparisi, E. (2009). (Eds.) *Multimodal metaphor*. Berlin: de Gruyter.

Foroni, F., & Semin, G. R. (2009). Language that puts you in touch with your bodily feelings the multimodal responsiveness of affective expressions. *Psychological Science*, *20*, 974-980.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression*. Thousand Oaks, CA: Sage.

Fraley, C., Raftery, A. E., Murphy, T. B., & Scrucca, L. (2012). mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation Technical Report No. 597, Department of Statistics, University of Washington.

Fryer, L., Freeman, J., & Pring, L. (2014). Touching words is not enough: How visual experience influences haptic–auditory associations in the "Bouba–Kiki" effect. *Cognition*, *132*, 164-173.

Gallace, A., Boschin, E., & Spence, C. (2011). On the taste of "Bouba" and "Kiki": An exploration of word–food associations in neurologically normal participants. *Cognitive Neuroscience*, *2*, 34-46.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in reason and language. *Cognitive Neuropsychology*, *22*, 455-479.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Cambridge, MA: MIT Press.

Gawel, R., Oberholster, A., & Francis, I. L. (2000). A 'Mouth-feel Wheel': terminology for communicating the mouth-feel characteristics of red wine. *Australian Journal of Grape and Wine Research*, *6*, 203-207.

Geeraerts, D. (1993). Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics*, *4*, 223-272.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. In *Technical report, Department of Statistics, Columbia University*. [www.stat. columbia.edu/~gelman/research/unpublished/p_hacking.pdf , Retrieved July 31, 2017]

Gentleman, R., & Lang, D. (2007). Statistical analyses and reproducible research. *Journal of Computational and Graphical Statistics*, *16*, 1-23.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*, 256-281.

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge, UK: Cambridge University Press.

Gibbs, R. W. (2005). *Embodiment and cognitive science*. Cambridge, UK: Cambridge University Press.

Gibbs, R. W. (2006). Metaphor interpretation as embodied simulation. *Mind & Language*, *21*, 434-458.

Gibbs, R. W. (2007). Why cognitive linguists should care more about empirical methods. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson & M. J. Spivey (Eds.), *Methods in cognitive linguistics* (pp. 2-18). Amsterdam: John Benjamins.

Gibbs, R. W., & Matlock, T. (2008). Metaphor, imagination, and simulation: Psycholinguistic evidence. In Gibbs, R. W. (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 161-176). Cambridge, UK: Cambridge University Press.

Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, *28*, 88-124.

Gilad, Y., Man, O., Pääbo, S., & Lancet, D. (2003). Human specific loss of olfactory receptor genes. *Proceedings of the National Academy of Sciences*, *100*, 3324-3327.

Gilad, Y., Wiebe, V., Przeworski, M., Lancet, D., & Pääbo, S. (2004). Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLOS Biology*, *2*, e5.

Givón, T. (2001). *Syntax: A functional-typological introduction, Vol. 1*. Amsterdam-Philadelphia: Benjamins.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, *20*, 1-55.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*, 558-565.

Glenberg, A.M., & Robertson, D.A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*, 1-26.

Glenberg, A.M., & Robertson, D.A. (2000). Symbol grounding and meaning: A comparison of high dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*, 379-401.

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006a). Perceptual knowledge retrieval activates sensory brain regions. *The Journal of Neuroscience*, 26, 4917-4921.

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006b). Distinct and common cortical activations for multimodal semantic categories. *Cognitive, Affective, & Behavioral Neuroscience*, *6*, 214-222.

Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, *23*, 959-978.

González, J., Barros-Loscertales, A., Pulvermüller, F., Meseguer, V., Sanjuán, A., Belloch, V., & Ávila, C. (2006). Reading cinnamon activates olfactory brain regions. *Neuroimage*, *32*, 906-912.

Gonzalez-Marquez, M., Mittelberg, I., Coulson, S. & Spivey, M. J. (2007). Introduction. The many faces of research in Cognitive Linguistics. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson & M. J. Spivey (Eds.), *Methods in Cognitive Linguistics* (pp. xi-xxi). Amsterdam: John Benjamins.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, *18*, 694-698.

Gori, M., Sandini, G., Martinoli, C., & Burr, D. (2010). Poor haptic orientation discrimination in nonsighted children may reflect disruption of cross-sensory calibration. *Current Biology*, *20*, 223-225.

Grady, J. (1997). THEORIES ARE BUILDINGS revisited. *Cognitive Linguistics*, *8*, 267-290.

Grady, J. (1999). A typology of motivation for conceptual metaphor: correlation vs. resemblance. In R. Gibbs & G. Steen (Eds.), *Metaphor in cognitive linguistics* (pp. 79-100). Amsterdam: John Benjamins.

Grady, J. (2005). Primary metaphors as inputs to conceptual integration. *Journal of Pragmatics*, *37*, 1595-1614.

Grady, J., Oakley, T., & Coulson, S. (1999). Blending and metaphor. In R. W. Gibbs, & G. J. Steen (Eds.), *Metaphor in cognitive linguistics* (pp. 101-124). Amsterdam: John Benjamins.

Graumann, A. (2007). Color names and dynamic imagery. In M. Plümacher, & P. Holz (Eds.), *Speaking of colors and odors* (pp. 129-140). Amsterdam: John Benjamins.

Gries, S. T., & Divjak, D. S. (2010). Quantitative approaches in usage-based cognitive semantics: myths, erroneous assumptions, and a proposal. In D. Glynn & K. Fischer (Eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches* (pp. 333-353). Berlin: de Gruyter.

Gries, S. T., Hampe, B., & Schönefeld, D. (2005). Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics*, *16*, 635-676.

Grondelaers, S., Geeraerts, D. & Speelman, D. (2007). A case for a cognitive corpus linguistics. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson & M. J. Spivey (Eds.), *Methods in Cognitive Linguistics* (pp. 149-169). Amsterdam: John Benjamins.

Grossenbacher, P. G., & Lovelace, C. T. (2001). Mechanisms of synesthesia: Cognitive and physiological constraints. *Trends in Cognitive Sciences*, *5*, 36-41.

Grote, K. (2013). “Modality relativity”: The influence of sign language and spoken language on conceptual categorization. PhD thesis, RWTH Aachen.

Guest, S., Catmur, C., Lloyd, D., & Spence, C. (2002). Audiotactile interactions in roughness perception. *Experimental Brain Research*, *146*, 161-171.

Haenny, P. E., Maunsell, J. H. R., & Schiller, P. H. (1988). State dependent activity in monkey visual cortex II: Retinal and extraretinal factors in V4. *Experimental Brain Research*, *69*, 245-259.

Hagen, M. C., Franzén, O., McGlone, F., Essick, G., Dancer, C., & Pardo, J. V. (2002). Tactile motion activates the human middle temporal/V5 (MT/V5) complex. *European Journal of Neuroscience*, *16*, 957-964.

Halgren, E. (1992). Emotional neurophysiology of the amygdala within the context of human cognition. In J. P. Aggleton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory and mental dysfunction* (pp. 191-228). New York: Wiley-Liss.

Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, *98*, 22-44.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, *13*, 70-84.

Haspelmath, M. (1999). Why is grammaticalization irreversible?. *Linguistics*, *37*, 1043-1068.

Hassemer, J., & Winter, B. (2018). Decoding gestural iconicity. *Cognitive Science*.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301-307.

Havas, D.A., Glenberg, A.M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review*, *14*, 436–441.

Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (2010). Cosmetic use of botulinum toxin-A affects processing of emotional language. *Psychological Science*, *21*, 895-900.

Havas, D. A., & Matheson, J. (2013). The functional role of the periphery in emotional language comprehension. *Frontiers in Psychology*, *4*, 294.

Havlicek, J., & Roberts, S. C. (2009). MHC-correlated mate choice in humans: a review. *Psychoeuroendocrinology*, *34*, 497-512.

Hay, J. C., & Pick, H. L. (1966). Visual and proprioceptive adaptation to optical displacement of the visual stimulus. *Journal of Experimental Psychology*, *71*, 150-158.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Healy, K. (2017). Fuck nuance. *Sociological Theory*, *35*, 118-127.

Heisenberg, W. (1958). *Physics and philosophy: The revolution in modern science. Lectures delivered at University of St. Andrews, Scotland, Winter 1955-56*. New York: Harper & Row.

Hermans, D., & Baeyens, F. (2002). Acquisition and activation of odor hedonics in everyday situations: Conditioning and priming studies. In C. Rouby, B. Schaal, D. Dubois, R. Gervais, & A. Holley (Eds.), *Olfaction, taste, and cognition* (pp. 119-139). Cambridge, UK: Cambridge University Press.

Herz, R. S. (2002). Influences of odors on mood and affective cognition. In C. Rouby, B. Schaal, D. Dubois, R. Gervais, & A. Holley (Eds.), *Olfaction, taste, and cognition* (pp. 160-177). Cambridge, UK: Cambridge University Press.

Herz, R. S. (2004). A naturalistic analysis of autobiographical memories triggered by olfactory visual and auditory stimuli. *Chemical Senses*, *29*, 217-24.

Herz, R. (2007). *The scent of desire: Discovering our enigmatic sense of smell*. New York: Harper Collins.

Herz, R. S., & Engen, T. (1996). Odor memory: Review and analysis. *Psychonomic Bulletin & Review*, *3*, 300-313.

Herz, R. S., & Schooler, J. W. (2002). A naturalistic study of autobiographical memories evoked by olfactory and visual cues: Testing the Proustian hypothesis. *American Journal of Psychology*, *115*, 21-32.

Hidaka, S., & Shimoda, K. (2014). Investigation of the effects of color on judgments of sweetness using a taste adaptation method. *Multisensory Research*, *27*, 189-205.

Hinton, L., Nichols, J., & Ohala, J. (1994). Introduction: sound-symbolic processes. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 1-12). Cambridge, UK: Cambridge University Press.

Hirata, S., Ukita, J., & Kita, S. (2011). Implicit phonetic symbolism in voicing of consonants and visual lightness using Garner's speeded classification task. *Perceptual Motor Skills*, *113*, 929-940.

Hockett, C. F. (1982 [1960]). The origin of speech. *Scientific American*, 203, 88–111. Reprinted in: W. S-Y Wang. (1982), *Human communication: Language and its psychobiological bases* (pp. 4–12). San Francisco: W. H. Freeman.

Howes, D. (1991) (Ed.). *The varieties of sensory experience: A sourcebook in the anthropology of the senses*. Toronto: University of Toronto Press.

Howes, D. (2002). Nose-wise: Olfactory metaphors in mind. In C. Rouby, B. Schaal, D. Dubois, R. Gervais, & A. Holley (Eds.), *Olfaction, taste, and cognition* (pp. 67-81). Cambridge, UK: Cambridge University Press.

Howes, D. (2006). Cross-talk between the senses. *The Senses & Society*, *1*, 381-390.

Huisman, J. L., & Majid, A. (2018). Psycholinguistic variables matter in odor naming. *Memory & Cognition*, *46*, 577-588.

Hultén, B. (2015). *Sensory marketing: Theoretical and empirical grounds*. London: Routledge.

Hultén, B., Broweus, N. & van Dijk, M. (2009). *Sensory marketing*. Houndmills: Palgrave.

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.

Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, *12*, 249-268.

Hunston, S. (2010). *Corpus approaches to evaluation: Phraseology and evaluative language*. London: Routledge.

Huumo, T. (2010). Is perception a directional relationship? On directionality and its motivation in Finnish expressions of sensory perception. *Linguistics*, *48*, 49-97.

Ibarretxe-Antuñano, I. (2008). Vision metaphors for the intellect: Are they really cross-linguistic?. *Atlantis*, *30*, 15-33.

IJzerman, H., & Koole, S. L. (2011). From perceptual rags to metaphoric riches—Bodily, social, and cultural constraints on sociocognitive metaphors: Comment on Landau, Meier, and Keefer (2010). *Psychological Bulletin*, *137*, 355-361.

IJzerman, H., & Semin, G. R. (2009). The thermometer of social relations mapping social proximity on temperature. *Psychological Science*, *20*, 1214-1220.

James, W. (1891 [1890]). *The principles of psychology*. Cambridge, MA: Harvard University Press.

Jastrzembski, J. E., & Stanners, R. F. (1975). Multiple word meanings and lexical search speed. *Journal of Verbal Learning and Verbal Behavior*, *14*, 534-537.

Johnson-Laird, P. N., & Quinn, J. G. (1976). To define true meaning. *Nature*, *264*, 635-636.

Jorgensen, J. C. (1990). The psychological reality of word senses. *Journal of Psycholinguistic Research*, 19, 167-190.

Jousmäki, V., & Hari, R. (1998). Parchment-skin illusion: sound-biased touch. *Current Biology*, *8*, R190-R191.

Judd, D. B., & Wyszecki, G. (1975). *Color in business, science and industry*. New York: Wiley-Interscience.

Julius, D., & Basbaum, A. I. (2001). Molecular mechanisms of nociception. *Nature*, *413*, 203-210.

Jurafsky, D. (2014). *The language of food*. New York: W. W. Norton.

Karns, C. M., & Knight, R. T. (2009). Intermodal auditory, visual, and tactile attention modulates early stages of neural processing. *Journal of Cognitive Neuroscience*, *21*, 669-683.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, *94*, B79-B89.

Kaschak, M. P., Zwaan, R. A., Aveyard, M., & Yaxley, R. H. (2006). Perception of auditory motion affects language processing. *Cognitive Science*, *30*, 733-744.

Katz, A. N., & Al-Azary, H. (2017). Principles that promote bidirectionality in verbal metaphor. *Poetics Today*, *38*, 35-59.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Kendon, A. (2014). Semiotic diversity in utterance production and the concept of 'language'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*, 20130293.

Keverne, E. B. (1999). The vomeronasal organ. *Science*, *286*, 716-720.

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience*, *28*, 12224-12230.

Kiritchenko, S., Zhu, X., & Mohammad, S. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, *50*, 723-762.

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics*, *35*, 379-415.

Klatzky, R. L., Lederman, S. J., & Reed, C. (1987). There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. *Journal of Experimental Psychology: General*, *116*, 356-369.

Köhler, R. (1986). *Zur Linguistischen Synergetik: Struktur und Dynamik der Lexik*. Bochum: Brockmeyer.

Köhler, W. (1929). *Gestalt psychology*. New York: Liveright.

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & Van der Schoot, M. (2016). Size does matter: Implied object size is mentally simulated during language comprehension. *Discourse Processes*, *54*, 493-503.

Koptjevskaja-Tamm, M. (2015) (Ed.) *The linguistics of temperature*. Amsterdam: John Benjamins.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 47-60.

Kosslyn, S. M., Seger, C., Pani, J. R., & Hillger, L. A. (1990). When is imagery used in everyday life? A diary study. *Journal of Mental Imagery*, *14*, 131-152.

Köster, E. P. (2002). The specific characeristics of the sense of smell. In C. Rouby, B. Schaal, D. Dubois, R. Gervais, & A. Holley (Eds.), *Olfaction, taste, and cognition* (pp. 27-43). Cambridge, UK: Cambridge University Press.

Kousta, S. T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112, 473-481.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General*, 140, 14-34.

Kövecses, Z. (2002). *Metaphor: A practical introduction*. Oxford: Oxford University Press.

Kövecses, Z. (2013). The metaphor–metonymy relationship: Correlation metaphors are based on metonymy. *Metaphor and Symbol*, *28*, 75-88.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*, 19-28.

Krifka, M. (2010). A note on the asymmetry in the hedonic implicatures of olfactory and gustatory terms. In S. Fuchs, P. Hoole, C. Mooshammer & M. Zygis (Eds.), *Between the regular and the particular in speech and language* (pp. 235-245). Frankfurt am Main: Peter Lang.

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 279-294.

Kuperman, V. (2015). Virtual experiments in megastudies: A case study of language and emotion. *The Quarterly Journal of Experimental Psychology*, 68, 1693-1710.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*, 1065-1081.

Lacey, S., Stilla, R., & Sathian, K. (2012). Metaphorically feeling: comprehending textural metaphors activates somatosensory cortex. *Brain and Language*, *120*, 416-421.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lakoff, G. (1990). The invariance hypothesis: is abstract reason based on image-schemas?. *Cognitive Linguistics*, *1*, 39-74.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. New York: Basic Books.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.

Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A metaphor-enriched social cognition. *Psychological Bulletin*, *136*, 1045-1067.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford, CA: Stanford university press.

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

Lascaratou, C. (2007). *The language of pain: Expression or description?*. Amsterdam: John Benjamins.

Lawless, H. T. (1984). Flavor description of white wine by "expert" and nonexpert wine consumers. *Journal of Food Science*, *49*, 120-123.

Lederman, S. J. (1979). Auditory texture perception. *Perception*, *8*, 93-103.

Lee, A. P. J. (2015). Lexical categories and conceptualization of olfaction in Amis. *Language and Cognition*, *7*, 321-350.

Lee, S. W., & Schwarz, N. (2012). Bidirectionality, mediation, and moderation of metaphorical effects: The embodiment of social suspicion and fishy smells. *Journal of Personality and Social Psychology*, *103*, 737-749.

Le Guérer, A. (2002). Olfaction and cognition: A philosophical and psychoanalytic view. In C. Rouby, B. Schaal, D. Dubois, R. Gervais, & A. Holley (Eds.), *Olfaction, taste, and cognition* (pp. 196-208). Cambridge, UK: Cambridge University Press.

Lehrer, A. (1978). Structures of the lexicon and transfer of meaning. *Lingua, 45*, 95-123.

Lehrer, A. (2009). *Wine and conversation* (2nd Ed.). Oxford: Oxford University Press.

Lemon, J. (2006). Plotrix: a package in the red light district of R. *R-News*, *6*, 8-12.

Levänen, S., Jousmäki, V., & Hari, R. (1998). Vibration-induced auditory-cortex activation in a congenitally deaf adult. *Current Biology*, *8*, 869-872.

Levinson, S. C. (2000). Yélî Dnye and the theory of basic color terms. Journal of *Linguistic Anthropology*, *10*, 3-55.

Levinson, S. C., & Majid, A. (2014). Differential ineffability and the senses. *Mind & Language*, *29*, 407-427.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Liddell, S. K. (2003). *Grammar, gesture and meaning in American sign language*. Cambridge, UK: Cambridge University Press.

Lindstrom, M. (2010). *Brand sense: Sensory secrets behind the stuff we buy* (2nd Ed.). London: Kogan Limited.

Littlemore, J. (2015). *Metonymy: Hidden shortcuts in language, thought and communication*. Cambridge, UK: Cambridge University Press.

Liu, B. (2012). *Sentiment analysis and opinion mining*. San Rafael, CA: Morgan & Claypool Publishers.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, *6*, 1246.

Lorig, T. S. (1999). On the similarity of odor and language perception. *Neuroscience & Biobehavioral Reviews*, *23*, 391-398.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis & T. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 157-176). Amsterdam: John Benjamins.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*, 273-302.

Louwerse, M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, *35*, 381-398.

Lu, J., & Goldin-Meadow, S. (2017). Combining categorical and gradient information in sign and spoken communication. *Presentation at the 11th International Symposium on Iconicity in Language and Literature*, Brighton, UK.

Lu, J. C., & Goldin-Meadow, S. (2018). Creating images with the stroke of a hand: Depiction of size and shape in sign language. *Frontiers in Psychology*, *9*, 1276.

Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge, UK: Cambridge University Press.

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*, 279-284.

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *373*, 20170137.

Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, *41*, 558-564.

Lynott, D., & Connell, L. (2013). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods*, *45*, 516-526.

Macpherson, F. (Ed.) (2011). *The senses: Classic and contemporary philosophical perspectives*. Oxford: Oxford University Press.

Maechler, M. (2015). diptest: Hartigan's dip test statistic for unimodality - corrected. R package version 0.75-7.

Mahmut, M. K., & Stevenson, R. J. (2015). Failure to obtain reinstatement of an olfactory representation. *Cognitive Science*, *39*, 1940-1949.

Majid, A. (2012). Current emotion research in the language sciences. *Emotion Review*, *4*, 432-443.

Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition*, *130*, 266-270.

Majid, A., Burenhult, N., Stensmyr, M., de Valk, J., & Hansson, B. S. (2018). Olfactory language and abstraction across cultures. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *373*, 20170139.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology*, *102*, 59-70.

Mannaert, L. N. H., Dijkstra, K., & Zwaan, R. A. (2017). Is color an integral part of a rich mental simulation?. *Memory & Cognition*, *45*, 974-982.

Marchand, H. (1959). Phonetic symbolism in English word formations. *Indogermanische Forschungen*, *64*, 146-168.

Marchand, H. (1960). *The categories and types of present-day English word formation*. Tuscaloosa, AL: University of Alabama Press.

Markman, A. B. (2008). Pluralism, relativism and the proper use of theories. *Journal of Experimental and Theoretical Artificial Intelligence*, *20*, 247-250.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology*, *87,* 173-188.

Marks, L. E. (1975). On colored-hearing synesthesia: cross-modal translations of sensory dimensions. *Psychological Bulletin*, *82*, 303-331.

Marks, L. E. (1978). *The unity of the senses: Interrelations among the modalities*. New York: Academic Press.

Marks, L. E. (1982a). Bright sneezes and dark coughs, loud sunlight and soft moonlight. *Journal of Experimental Psychology, Human Perception and Performance*, *8*, 177-193.

Marks, L. E. (1982b). Synesthetic perception and poetic metaphor. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 15-23.

Marks, L. E., & Stevens, J. C. (1966). Individual brightness functions. *Perception & Psychophysics*, *1*, 17-24.

Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science*, *10*, 61-65.

Matlock, T. (1989). Metaphor and the grammaticalization of evidentials. In *Proceedings of the 15th Annual Meeting of the Berkeley Linguistics Society* (pp. 215-225). Berkeley: Berkeley Linguistics Society.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science*, *9*, 316-322.

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton, FL: CRC Press.

McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.

Meir, I. (2010). Iconicity and metaphor: Constraints on metaphorical extension of iconic forms. *Language*, *86*, 865-896.

Mendelson, D. (1984). Synaesthetic metaphor in the work of Isaak Babel. *Russian Literature*, *15*, 347-362.

Mesirov, J. P. (2010). Computer science. Accessible reproducible research. *Science*, *327*, 5964.

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs language affects low-level visual perception. *Psychological Science*, *18*, 1007-1013.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*, 788-804.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, *38*, 39-41.

Miller, G.A. & Charles W. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*, 1-28.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

Miller, I. J., & Reedy, F. E. (1990). Variations in human taste bud density and taste intensity perception. *Physiology & Behavior*, *47*, 1213-1219.

Mitchell, S. D. (2004). Why integrative pluralism? *E:CO*, *6*, 81-91.

Mitchell, S. D., & Dietrich, M. R. (2006). Integration without unification: An argument for pluralism in the biological sciences. *The American Naturalist*, *168*, 73-79.

McBurney, D. H. (1986). Taste, smell, and flavor terminology: Taking the confusion out of the fusion. In H. L. Meiselman, & R. S. Rivkin (Eds.), *Clinical measurement of taste and smell* (pp. 117-125). New York: Macmillan.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.

Mohammad, S. M., & Kiritchenko, S. (2015). Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, *31*, 301-326.

Mojet, J., Köster, E. P., & Prinz, J. F. (2005). Do tastants have a smell?. *Chemical Senses*, *30*, 9-21.

Moore, T., & Carling, C. (1988). *The limitations of language*. London: Macmillan.

Moos, A., Simmons, D., Simner, J., & Smith, R. (2013). Color and texture associations in voice-induced synesthesia. *Frontiers in Psychology*, *4*, 568.

Møller, A. (2012). *Sensory systems: Anatomy and physiology* (2nd Ed.). Richardson: A. R. Møller Publishing.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research*, *17*, 154-163.

Morley, J., & Partington, A. (2009). A few frequently asked questions about semantic—or evaluative—prosody. *International Journal of Corpus Linguistics*, *14*, 139-158.

Morrot, G., Brochet, F., & Dubourdieu, D. (2001). The color of odors. *Brain and Language*, *79*, 309-320.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J. & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 0021.

Murphy, G. L. (1996). On metaphoric representation. *Cognition*, *60*, 173-204.

Murphy, G. L. (1997). Reasons to doubt the present evidence for metaphoric representation. *Cognition*, *62*, 99-108.

Murphy, M. L. (2010). *Lexical meaning*. Cambridge, UK: Cambridge University Press.

Nagata, H. (1988). The relativity of linguistic intuition: The effect of repetition on grammaticality judgments. *Journal of Psycholinguistic Research*, *17*, 1-17.

Nagata, H. (1989). Effect of repetition on grammaticality judgments under objective and subjective self-awareness conditions. *Journal of Psycholinguistic Research*, *18*, 255-269.

Nagata, H. (1992). Anchoring effects in judging grammaticality of sentences. *Perceptual and Motor Skills*, *75*, 159-164.

Nakagawa, H. (2012). The importance of TASTE verbs in some Khoe languages. *Linguistics*, *50*, 395-420.

Nakamura, T., Sakamoto, M., & Utsumi, A. (2010). The role of event knowledge in comprehending synesthetic metaphors. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 1898-1903). Austin, TX: Cognitive Science Society.

Ngo, M. K., Misra, R., & Spence, C. (2011). Assessing the shapes and speech sounds that people associate with chocolate samples varying in cocoa content. *Food Quality and Preference*, *22*, 567-572.

Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, *65*, 115-124.

Nielsen, A., & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition*, *4*, 115-125.

Nielsen, A., & Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology*, *67*, 153-163.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231-259.

Noble, A. C., Arnold, R. A., Buechsenstein, J., Leach, E. J., Schmidt, J. O., & Stern, P. M. (1987). Modification of a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture*, *38*, 143-146.

Novich, S., Cheng, S., & Eagleman, D. M. (2011). Is synaesthesia one condition or many? A large-scale analysis reveals subgroups. *Journal of Neuropsychology*, *5*, 353-371.

O'Callaghan, C. (2009). Sounds and events. In M. Nudds & C. O'Callaghan (Eds.), *Sounds & perception: New philosophical essays* (pp. 26-49). Oxford: Oxford University Press.

O'Callaghan, C. (2015). Not all perceptual experience is modality specific. In D. Stokes, M. Matthen & S. Biggs (Eds.), *Perception and its modalities* (pp. 133-165). Oxford: Oxford University Press.

O'Callaghan, C. & Nudds, M. (2009). Introduction: The philosophy of sounds and auditory perception. In M. Nudds & C. O'Callaghan (Eds.), *Sounds & perception: New philosophical essays* (pp. 1-25). Oxford: Oxford University Press.

O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Mahwah: Lawrence Erlbaum.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*, 1-16.

Ohala, J. J. (1994). The frequency code underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 325-347). Cambridge, UK: Cambridge University Press.

Olofsson, J. K., & Gottfried, J. A. (2015). The muted sense: neurocognitive limitations of olfactory language. *Trends in Cognitive Sciences*, *19*, 314-321.

O'Malley, G. (1957). Literary synesthesia. *The Journal of Aesthetics and Art Criticism*, *15*, 391-411.

Ortony, A. (1975). Why metaphors are necessary and not just nice. *Educational Theory*, 25, 45-53.

Ortiz, M. J. (2011). Primary metaphors and monomodal visual metaphors. *Journal of Pragmatics*, *43*, 1568-1580.

Osaka, N., Osaka, M., Morishita, M., Kondo, H., & Fukuyama, H. (2004). A word expressing affective pain activates the anterior cingulate cortex in the human brain: an fMRI study. *Behavioural Brain Research*, *153*, 123-127.

Osgood, C. (1963). Language universals and psycholinguistics. In J. Greenberg (Ed.), *Universals of language* (pp. 236-254). Cambridge, Mass: MIT Press.

Osgood, C. E. (1981). The cognitive dynamics of synesthesia and metaphor. *Review of Research in Visual Arts Education, 7*, 56-80.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery and meaningfulness values for 925 words. *Journal of Experimental Psychology Monograph Supplement*, *76*.

Palmer, S. E., Schloss, K. B., Xu, Z., & Prado-León, L. R. (2013). Music–color associations are mediated by emotion. *Proceedings of the National Academy of Sciences*, *110*, 8836-8841.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics* (article 271). Association for Computational Linguistics.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*, 1-135.

Papesh, M. H. (2015). Just out of reach: On the reliability of the action-sentence compatibility effect. *Journal of Experimental Psychology: General*, *144*, e116-e141.

Paradis, C. (2000). Reinforcing adjectives: A cognitive semantic approach on grammaticalization. In R. Bermudez-Otero, D. Denison, R. M. Hogg, & C. B. McCully (Eds)., *Generative theory and corpus studies* (pp. 233-258). Berlin: de Gruyter.

Paradis, C. (2005). Ontologies and construals in lexical semantics. *Axiomathes*, *15*, 541-573.

Paradis, C., & Eeg-Olofsson, M. (2013). Describing sensory experience: The genre of wine reviews. *Metaphor and Symbol*, *28*, 22-40.

Partington, A. (2004). "Utterly content in each other's company": Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, *9*, 131-156.

Pasnau, R. (1999). What is sound?. *The Philosophical Quarterly*, *49*, 309-324.

Patel, A. D., & Iversen, J. R. (2003). Acoustic and perceptual comparison of speech and drum sounds in the north indian tabla tradition: An empirical study of sound symbolism. In M. J. Solé, D. Recansens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 925-928). Barcelona.

Pecher, D., van Dantzig, S., Zwaan, R. A., & Zeelenberg, R. (2009). Language comprehenders retain implied shape and orientation of objects. *The Quarterly Journal of Experimental Psychology*, *62*, 1108-1114.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, *14*, 119-124.

Peli, E., Fine, E. M., & Labianca, A. T. (1996). Evaluating visual information provided by audio description. *Journal of Visual Impairment and Blindness*, *90*, 378-385.

Peng, R. D. (2011). Reproducible research in computational science. *Science*, *334*, 1226-1227.

Pennycook, G., & Thompson, V. A. (2017). Base-rate neglect. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (pp. 44-61). London: Routledge.

Pérez-Sobrino, P. (2016). Multimodal metaphor and metonymy in advertising: A corpus-based account. *Metaphor and Symbol*, *31*, 73-90.

Pérez-Sobrino, P. (2017). *Multimodal metaphor and metonymy in advertising*. Amsterdam: John Benjamins.

Pérez-Sobrino, P., & Julich, N. (2014). Let's talk music: A corpus-based account of musical motion. *Metaphor and Symbol*, *29*, 298-315.

Perlman, M. (2010). Talking fast: The use of speech rate as iconic gesture. In F. Perrill, V. Tobin, & M. Turner (Eds.), *Meaning, form, and body* (pp. 245-262). Stanford: CSLI Publications.

Perlman, M., & Cain, A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, *14*, 321-351.

Perlman, M., Clark, N., & Johansson Falck, M. (2014). Iconic prosody in story reading. *Cognitive Science*, *39*, 1348-1368.

Perlman, M., Little, H., Thompson, B., & Thompson, R. L. (2018). Iconicity in signed and spoken vocabulary: A comparison between American Sign Language, British Sign Language, English, and Spanish. *Frontiers in Psychology*, *9*, 1433.

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLOS ONE*, *10*, e0137147.

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*, *21*, e12572.

Petersen, W., Fleischhauer, J., Beseoglu, H., & Bücker, P. (2008). A frame-based analysis of synaesthetic metaphors. *Baltic International Yearbook of Cognition, Logic and Communication*, *3*, 1-22.

Pfeiffer, J. C., Hollowood, T. A., Hort, J., & Taylor, A. J. (2005). Temporal synchrony and integration of sub-threshold taste and smell signals. *Chemical Senses*, *30*, 539-545.

Phillips, M. L., & Heining, M. (2002). Neural correlates of emotion perception: From faces to taste. In C. Rouby, B. Schaal, D. Dubois, R. Gervais, & A. Holley (Eds.), *Olfaction, taste, and cognition* (pp. 196-208). Cambridge, UK: Cambridge University Press.

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, *6*, 203-205.

Plümacher, M., & Holz, P. (2007). Speaking of colors and odors. In M. Plümacher, & P. Holz (Eds.), *Speaking of colors and odors* (pp. 1-17). Amsterdam: John Benjamins.

Plümacher, M. (2007). Speaking of colors and odors. In M. Plümacher, & P. Holz (Eds.), *Speaking of colors and odors* (pp. 61-84). Amsterdam: John Benjamins.

Pomp, J., Bestgen, A. K., Schulze, P., Müller, C. J., Citron, F. M., Suchan, B., & Kuchinke, L. (2018). Lexical olfaction recruits olfactory orbitofrontal cortex in metaphorical and literal contexts. *Brain and Language*, *179*, 11-21.

Popova, Y. B. (2003). 'The fool sees with his nose': metaphoric mappings in the sense of smell in Patrick Süskind's Perfume. *Language and Literature*, *12*, 135-151.

Popova, Y. (2005). Image schemas and verbal synaesthesia. In B. Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics* (pp. 395-420). Berlin: de Gruyter.

Porat, R., & Shen, Y. (2017). Metaphor: The journey from bidirectionality to unidirectionality. *Poetics Today*, *38*, 123-140.

Porcello, T. (2004). Speaking of sound: Language and the professionalization of sound-recording engineers. *Social Studies of Science, 34*, 733-758.

Pragglejaz Group, P. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, *22*, 1-39.

Prather, S. C., Votaw, J. R., & Sathian, K. (2004). Task-specific recruitment of dorsal and ventral visual areas during tactile perception. *Neuropsychologia*, *42*, 1079-1087.

Price, J. L. (1987). The central and accessory olfactory systems. In T. E. Finger & W. L. Silver (Eds.), *Neurobiology of taste and smell* (pp. 179-204). New York: Wiley.

Pullum, G. K. (2007). Ungrammaticality, rarity, and corpus use. *Corpus Linguistics and Linguistic Theory*, *3*, 33-47.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*, 576-582.

Rakova, M. (2003). *The extent of the literal: Metaphor, polysemy and theories of concepts*. New York: Palgrave Macmillan.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia—a window into perception, thought and language. *Journal of Consciousness Studies*, *8*, 3-34.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PLOS ONE*, *11*, e0151138.

Richardson, M. P., Strange, B. A., & Dolan, R. J. (2004). Encoding of emotional memories depends on amygdala and hippocampus and their interactions. *Nature Neuroscience*, 7, 278-285.

Richter, M., Eck, J., Straube, T., Miltner, W. H., & Weiss, T. (2010). Do words hurt? Brain activation during the processing of pain-related words. *Pain*, *148*, 198-205.

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52-66.

Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, *143*, 594-596.

Rolls, E. (2008). Functions of the orbitofrontal and pregenual cingulate cortex in taste, olfaction, appetite and emotion. *Acta Physiologica Hungarica*, *95*, 131-164.

Ronga, I. (2016). Taste synaesthesias: Linguistic features and neurophysiological bases. In E. Gola & F. Ervas (Eds.), *Metaphor and communication* (pp. 47-60). Amsterdam: John Benjamins.

Ronga, I., Bazzanella, C., Rossi, F., & Iannetti, G. (2012). Linguistic synaesthesia, perceptual synaesthesia, and the interaction between multiple sensory modalities. *Pragmatics & Cognition*, *20*, 135-167.

Rouby, C., & Bensafi, M. (2002). Is there a hedonic dimension to odors? In C. Rouby, B. Schaal, D. Dubois, R. Gervais, & A. Holley (Eds.), *Olfaction, Taste, and Cognition* (pp. 140-159). Cambridge, UK: Cambridge University Press.

Royet, J. P., Plailly, J., Delon-Martin, C., Kareken, D. A., & Segebarth, C. (2003). fMRI of emotional responses to odors: influence of hedonic valence and judgment, handedness, and gender. *Neuroimage*, *20*, 713-728.

Royet, J. P., Zald, D., Versace, R., Costes, N., Lavenne, F., Koenig, O., & Gervais, R. (2000). Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: a positron emission tomography study. *The Journal of Neuroscience*, *20*, 7752-7759.

Sadamitsu, M. (2004). Synaesthesia re-examined: an alternative treatment of smell related concepts. *Osaka University Papers in English Linguistics*, *8*, 109-125.

Sakamoto, M., & Utsumi, A. (2014). Adjective metaphors evoke negative meanings. *PLOS ONE*, *9*, e89008.

Sakamoto, M., & Watanabe, J. (2016). Cross-modal associations between sounds and drink tastes/textures: a study with spontaneous production of sound-symbolic words. *Chemical Senses*, *41*, 197-203.

Sakamoto, M., & Watanabe, J. (2018). Bouba/Kiki in touch: Associations between tactile perceptual qualities and Japanese phonemes. *Frontiers in Psychology*, *9*, 295.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*, 225-239.

San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., Dirksmeyer, T., Enfield, N., Floyd, S., Hammond, H., Rossi, G., Tufvesson, S., van Putten, S., & Majid, A. (2014). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics*, *26*, 31-60.

San Roque, L., Kendrick, K. H., Norcliffe, E., & Majid, A. (2018). Universal meaning extensions of perception verbs are grounded in interaction. *Cognitive Linguistics*, *29*, 371-406.

Sathian, K., & Zangaladze, A. (2002). Feeling with the mind's eye: contribution of visual cortex to tactile perception. *Behavioural Brain Research*, *135*, 127-132.

Sathian, K., Zangaladze, A., Hoffman, J. M., & Grafton, S. T. (1997). Feeling with the mind's eye. *Neuroreport*, *8*, 3877-3881.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Szarkowska, A. (2011). Text-to-speech audio description: towards wider availability of AD. *Journal of Specialised Translation*, *15*, 81-98.

de Saussure, F. (1959) [1916]. *Course in general linguistics*. New York: The philosophical library.

Schiffman, S., Robinson, D. E., & Erickson, R. P. (1977). Multidimensional scaling of odorants: Examination of psychological and physicochemical dimensions. *Chemical Senses*, *2*, 375-390.

Schmeidler, E. & Kirchner, C. (2001). Adding audio description: Does it make a difference?. *Journal of Visual Impairment & Blindness*, *95*, 197-212.

Schmidtke, D. S., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in Psychology*, *5*, 80.

Schott, D. G. (2004). Communicating the experience of pain: the role of analogy. *Pain*, *108*, 209-212.

Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *Journal of Neurophysiology*, *85*, 1322-1327.

Schürmann, M., Caetano, G., Jousmäki, V., & Hari, R. (2004). Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. *The Journal of the Acoustical Society of America*, *115*, 830-832.

Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Searle, J. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press.

Semino, E. (2010). Descriptions of pain, metaphor, and embodied simulation. *Metaphor and Symbol*, *25*, 205-226.

Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain*, 115, 15-36.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, *14*, 147-152.

Shaoul, C., & Westbury, C. (2013). A reduced redundancy usenet corpus (2005-2011). Edmonton, AB: University of Alberta. Retrieved from

<http://www.psych.ualberta.ca/westburylab/downloads/usenetcorpus.download.html>.

Shepard, R. N., & Cooper, L. A. (1992). Representation of colors in the blind, color-blind, and normally sighted. *Psychological Science*, *3*, 97-104.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701-703.

Shen, Y. (1997). Cognitive constraints on poetic figures. *Cognitive Linguistics*, *8*, 33-71.

Shen, Y. (1998). How come silence is sweet but sweetness is not silent: a cognitive account of directionality in poetic synaesthesia. *Language and Literature*, *7*, 123-140.

Shen, Y. (2008). Metaphor and poetic figures. In R. W. Gibbs (Ed.), *Cambridge handbook of metaphor and thought* (pp. 295-307). Cambridge, UK: Cambridge University Press.

Shen, Y., & Aisenman, R. (2008). Heard melodies are sweet, but those unheard are sweeter: Synaesthetic metaphors and cognition. *Language and Literature*, *17*, 107-121.

Shen, Y., & Cohen, M. (1998). How come silence is sweet but sweetness is not silent: a cognitive account of directionality in poetic synaesthesia. *Language and Literature*, *7*, 123-140.

Shen, Y., & Gil, D. (2007). Sweet fragrances from Indonesia: A universal principle governing directionality in synaesthetic Metaphors. In W. van Peer, & J. Auracher (Eds.), *New beginnings in literary studies* (pp. 49-71). Newcastle, UK: Cambridge Scholars Publishing.

Shen, Y., & Gadir, O. (2009). How to interpret the music of caressing: Target and source assignment in synaesthetic genitive constructions. *Journal of Pragmatics*, *41*, 357-371.

Shen, Y., & Porat, R. (2017). Metaphorical directionality: The role of language. In B. Hampe (Ed.), *Metaphor: Embodied cognition and discourse* (pp. 62-81). Cambridge, UK: Cambridge University Press.

Shermer, D. Z., & Levitan, C. A. (2014). Red hot: The crossmodal effect of color intensity on perceived piquancy. *Multisensory Research*, *27*, 207-223.

Shibuya, Y., & Nozawa, H. (2003). Constraints on synaesthesia. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* (Vol. 29, No. 1, pp. 403-414). Berkeley: Berkeley Linguistics Society.

Shibuya, Y., Nozawa, H., & Kanamaru, T. (2007). Understanding synesthetic expressions: Vision and olfaction with the physiological = psychological model. In M. Plümacher, & P. Holz (Eds.), *Speaking of colors and odors* (pp. 203-226). Amsterdam: John Benjamins.

Shinohara, K., & Nakayama, A. (2011). Modalities and directions in synaesthetic metaphors in Japanese. *Cognitive Studies*, *18*, 491-507.

Shintel, H., & Nusbaum, H. C. (2007). The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. *Cognition*, *105*, 681-690.

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language*, *55*, 167-177.

Sidhu, D. M., & Pexman, P. M. (2017). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, 1-25.

Sidhu, D. M., & Pexman, P. M. (2018). Lonely sensational icons: semantic neighbourhood density, sensory experience and iconicity. *Language, Cognition and Neuroscience*, *33*, 25-31.

Silberzahn et al. (2017). Many analysts, one dataset: Making transparent how variations in analytical choices affect results. *Open Science Framework* [https://osf.io/qix4g/ , accessed May 29, 2017]

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359-1366.

Simner, J. (2012). Defining synaesthesia. *British Journal of Psychology*, *103*, 1-15.

Simner, J., Cuskley, C., & Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception*, *39*, 553-569.

Simner, J., Harrold, J., Creed, H., Monro, L., & Foulkes, L. (2008). Early detection of markers for synaesthesia in childhood populations. *Brain*, *132*, 57-64.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., Scott, K., & Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception*, *35*, 1024-1033.

Slobin, D. (1971). *Psycholinguistics*. Glenview, IL: Scott Foresman & Co.

Smeets, M. A. M., & Dijksterhuis, G. B. (2014). Smelly primes–when olfactory primes do or do not work. *Frontiers in Psychology*, *5*, 96.

Snefjella, B., & Kuperman, V. (2016). It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition*, *156*, 135-146.

Sorabji, R. (1971). Aristotle on demarcating the five senses. *The Philosophical Review*, *80*, 55-79.

Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Congruency effects between auditory and tactile motion: extending the phenomenon of cross-modal dynamic capture. *Cognitive, Affective, & Behavioral Neuroscience*, *4*, 208-217.

Suárez Toste, E. (2007). Metaphor inside the wine cellar: On the ubiquity of personification schemas in winespeak. *Metaphorik*, *12*, 53-64.

Speed, L. J., & Majid, A. (2017). Dutch modality exclusivity norms: Simulating perceptual modality in space. *Behavior Research Methods*, *49*, 1-15.

Speed, L. J., & Majid, A. (2018). An exception to mental simulation: No evidence for embodied odor language. *Cognitive Science*, *42*, 1146-1178.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustical Science and Technology*, *28*, 61-70.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*, 971-995.

Spence, C. (2012). Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism. *Journal of Consumer Psychology*, *22*, 37-54.

Spence, C. (2013). Multisensory flavour perception. *Current Biology*, *23*, R365-R369.

Spence, C. (2015). Multisensory flavor perception. *Cell*, *161*, 24-35.

Spence, C. (2015). Eating with our ears: Assessing the importance of the sounds of consumption to our perception and enjoyment of multisensory flavour experiences. *Flavour*, *4*, 3.

Spence, C., & Bayne, T. (2015). Is consciousness multisensory? In D. Stokes, M. Matthen & S. Biggs (Eds.), *Perception and its modalities* (pp. 95-132). Oxford: Oxford University Press.

Spence, C., Nicholls, M. E., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*, 330-336.

Spence, C., & Piqueras-Fiszman, B. (2014). *The perfect meal: The multisensory science of food and dining*. Chichester, UK: Wiley Blackwell.

Spence, C., Smith, B., & Auvray, M. (2015). Confusing tastes and flavours. In D. Stokes, M. Matthen, & S. Biggs (Eds.), *Perception and its modalities* (pp. 247-274). Oxford: Oxford University Press.

Spencer, N. J. (1973). Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguistic Research*, *2*, 83-98.

Spivey, M. (2007). *The continuity of mind*. Oxford: Oxford University Press.

Sprouse, J. & Hornstein, N. (2013). *Experimental syntax and island effects*. Cambridge, UK: Cambridge University Press.

Stadtlander, L. M., & Murdoch, L. D. (2000). Frequency of occurrence and rankings for touch-related adjectives. *Behavior Research Methods, 32*, 579-587.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*, 153-156.

Staniewski, P. (2017). Geschmack und smak - sprachliche Aspekte der gustatorischen Wahrnehmung im Deutschen und Polnischen. In J. Szczek & M. Kalasznik (Eds.), *Beiträge zur Fremdsprachenvermittlung* (pp. 223-248). Landau: Verlag Empirische Pädagogik.

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*. Amsterdam: John Benjamins.

Stevenson, R. J., Prescott, J., & Boakes, R. A. (1999). Confusing tastes and smells: how odours can influence the perception of sweet and sour tastes. *Chemical Senses*, *24*, 627-635.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. New York: Wiley.

Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences*, *54*, 407-411.

Stewart, D. (2010). *Semantic prosody: A critical evaluation*. London: Routledge.

Stoffregen, T. A., & Bardy, B. G. (2001). On specification and the senses. *Behavioral and Brain Sciences*, *24*, 195-213.

Stokes, D., & Biggs, S. (2015). The dominance of the visual. In D. Stokes, M. Matthen & S. Biggs (Eds.), *Perception and its modalities* (pp. 350-378). Oxford: Oxford University Press.

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, *54*, 768-777.

Strik Lievers, F. (2015). Synaesthesia: A corpus-based study of cross-modal directionality. In R. Caballero, & C. Paradis (Eds.), *Functions of language, sensory perceptions in language and cognition* (pp. 69-95). Amsterdam: John Benjamins.

Strik Lievers, F. (2016). Synaesthetic metaphors in translation. *Studi e Saggi Linguistici*, *54*, 43-70.

Strik Lievers, F. (2016). Figures and the senses: Towards a definition of synaesthesia. *Review of Cognitive Linguistics*, *15*, 83-101.

Strik Lievers, F., & Winter, B. (2018). Sensory language across lexical categories. *Lingua*, *204*, 45-61.

Stubbs, M. (2001). *Words and phrases*. Oxford: Blackwell.

Styles, S. J., & Gawne, L. (2017). When does maluma/takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter. *i-Perception*, *8*.

Suárez-Toste, E. (2013). One man's cheese is another man's music: Synaesthesia and the bridging of cultural differences in the language of sensory perception. In R. Caballero & J. Díaz-Vera (Eds.), *Sensuous Cognition* (pp. 169-191). Amsterdam: John Benjamins.

Sullivan, K., & Jiang, W. (2013). When my eyes are on you, do you touch my eyes? A reclassification of metaphors mapping from physical contact to perception. In T. Fuyin Li (Eds.), *Compendium of cognitive linguistics research volume 2* (pp. 189-200). Hauppauge, NY: Nova Science Publishers.

Sun, C., Koppel, K., & Chambers, E. (2014). An initial lexicon of sensory properties for nail polish. *International Journal of Cosmetic Science*, *36*, 262-272.

Suzuki, Y., Gyoba, J., & Sakamoto, S. (2008). Selective effects of auditory stimuli on tactile roughness perception. *Brain Research*, *1242*, 87-94.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge, UK: Cambridge University Press.

Synnott, A. (1991). Puzzling over the senses: From Plato to Marx. In D. Howes (Ed.), *The varieties of sensory experience: A sourcebook in the anthropology of the senses* (pp. 61-76). Toronto: University of Toronto Press.

Talavera, M., & Chambers, D. H. (2016). Flavor lexicon and characteristics of artisan goat cheese from the United States. *Journal of Sensory Studies*, *31*, 492-506.

Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, *2*, 325-347.

Taylor, J. R. (1995). *Linguistic categorization*. Oxford: Oxford University Press.

Teghtsoonian, M., & Teghtsoonian, R. (1965). Seen and felt length. *Psychonomic Science*, *3*, 465-466.

Tekiroğlu, S. S., Özbal, G., & Strapparava, C. (2014). A computational approach to generate a sensorial lexicon. In *Proceedings of the COLING 2014 Workshop* *on Cognitive Aspects of the Lexicon* (CogALex), August 2014, Dublin, Ireland.

Thagard, P. R. (1978). The best explanation: Criteria for theory choice. *The Journal of Philosophy*, *75*, 76-92.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, *106*, 4567-4570.

Thompson, G., & Hunston, S. (2000). Evaluation: An introduction. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 38-55). Oxford: Oxford University Press.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology*, *64*, 2392-2404.

Thoret, E., Aramaki, M., Kronland-Martinet, R., Velay, J. L., & Ystad, S. (2014). From sound to shape: auditory perception of drawing movements. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 983-994.

Thorndike, E. L. (1948). On the frequency of semantic changes in modern English. *The Journal of General Psychology*, *39*, 23-27.

Tomasino, B., Fink, G. R., Sparing, R., Dafotakis, M., & Weiss, P. H. (2008). Action verbs and the primary motor cortex: a comparative TMS study of silent reading, frequency judgments, and motor imagery. *Neuropsychologia*, *46*, 1915-1926.

Tracey, I. (2005). Nociceptive processing in the human brain. *Current Opinion in Neurobiology*, *15*, 478-487.

Troche, J., Crutch, S. J., & Reilly, J. (2017). Defining a conceptual topography of word concreteness: Clustering properties of emotion, sensation, and magnitude among 750 English words. *Frontiers in Psychology*, *8*, 1787.

Tsur, R. (2007). Issues in literary synaesthesia. *Style*, *41*, 30-51.

Tsur, R. (2008). *Toward a theory of cognitive poetics*. Brighton: Sussex Academic Press.

Tsur, R. (2012). *Playing by ear and the tip of the tongue: Precategorical information in poetry*. Amsterdam: John Benjamins.

Turatto, M., Galfano, G., Bridgeman, B., & Umiltà, C. (2004). Space-independent modality-driven attentional capture in auditory, tactile and visual systems. *Experimental Brain Research*, *155*, 301-310.

Turner, B. H., Mishkin, M., & Knapp, M. (1980). Organization of the amygdalopetal projections from modality-specific cortical association areas in the monkey. *Journal of Comparative Neurology*, *191*, 515-543.

Ullmann, S. (1945). Romanticism and synaesthesia: A comparative study of sense transfer in Keats and Byron. *Publications of the Modern Language Association of America*, *60*, 811-827.

Ullmann, S. (1959). *The principles of semantics*. Glasgow: Jackson, Son & Co.

Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language, Vol 2: Phonology* (pp. 525-568). Stanford, CA: Stanford University Press.

Urbanek, S. (2013). png: Read and write PNG images. R package version 0.1-7. https://CRAN.R-project.org/package=png

Usnadze, D. (1924). Ein experimenteller Beitrag zum Problem der psychologischen Grundlagen der Namengebung. *Psychologische Forschung*, *5*, 24-43.

Velasco-Sacristán, M., & Fuertes-Olivera, P. A. (2006). Olfactory and olfactory-mixed metaphors in print ads of perfume. *Annual Review of Cognitive Linguistics*, *4*, 217-252.

Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with S* (4th Ed.). New York: Springer.

Viberg, Å. (1983). The verbs of perception: a typological study. *Linguistics*, *21*, 123-162.

Viberg, Å. (1993). Crosslinguistic perspectives on lexical organization and lexical progression. In K. Hyltenstam, & Å. Viberg (Eds.), *Progression and regression in language: Sociocultural, neuropsychological and linguistic perspectives* (pp. 340–385). Cambridge, UK: Cambridge University Press.

Vickers, Z. M. (1984). Crispness and crunchiness – A difference in pitch? *Journal of Texture Studies*, *15*, 157-163.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*, 219-247.

Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods*, *40*, 1079-1087.

Volkow, N. D., Wang, G. J., & Baler, R. D. (2011). Reward, dopamine and the control of food intake: implications for obesity. *Trends in Cognitive Sciences*, *15*, 37-46.

Vukovic, N., Fardo, F., & Shtyrov, Y. (under review). When words burn – language processing differentially modulates pain perception in typical and chronic pain populations. *Language and Cognition*.

Vukovic, N., & Williams, J. N. (2014). Automatic perceptual simulation of first language meanings during second language sentence processing in bilinguals. *Acta Psychologica*, *145*, 98-103.

Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams Jr, R. B., Albohn, D. N., Allard, E. S., Benning, S. D., Blouin-Hudon, E.-M., Bulnes, L. C., Caldwell, T. L., Calin-Jageman, R. J., Capaldi, C. A., Carfagno, N. S., Chasten, K. T., Cleeremans, A., Connell, L., DeCicco, J. M., Dijkstra, K., Fischer, A. H., Foroni, F., Hess, U., Holmes, K. J., Jones, J. L. H., Klein, O., Koch, C., Korb, S., Lewinski, P., Liao, J. D., Lund, S., Lupianez, J., Lynott, D., Nance, C. N., Oosterwijk, S., Ozdoğru, A. A., Pacheco-Unguetti, A. P., Pearson, B., Powis, C., Riding, S., Roberts, T.-A., Rumiati, R. I., Senden, M., Shea-Shumsky, N. B., Sobocko, K., Soto, J. A., Steiner, T. G., Talarico, J. M., van Allen, Z. M., Vandekerckhove, M., Wainwright, B., Wayand, J. F., Zeelenberg, R., Zetzer, E. E., & Zwaan, R. A. (2016). Registered replication report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, *11*, 917-928.

Walker, E., & Cooperrider, K. (2016). The continuity of metaphor: Evidence from temporal gestures. *Cognitive Science*, *40*, 481-495.

Walsh, V. (2000). Neuropsychology: The touchy, feely side of vision. *Current Biology*, *10*, R34-R35.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*, 483-488.

Warriner, A.B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*, 1191-1207.

Warriner, A. B., & Kuperman, V. (2015). Affective biases in English are bi-dimensional. *Cognition and Emotion*, *29*, 1147-1167.

Waskul, D. D., Vannini, P., & Wilson, J. (2009). The aroma of recollection: Olfaction, nostalgia, and the shaping of the sensuous self. *The Senses and Society*, *4*, 5-22.

Wasow, T., & Arnold, J. (2005). Intuitions in linguistic argumentation. *Lingua*, *115*, 1481-1496.

Watanabe, J., Utsunomiya, Y., Tsukurimichi, H., & Sakamoto, M. (2012). Relationship between Phonemes and Tactile-emotional Evaluations in Japanese Sound Symbolic Words. In N. Miyake, D. Peebles, & R. P.

Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 2517-2522). Austin, TX: Cognitive Science Society.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*, 638-667.

Werning, M., Fleischhauer, J., & Beseoglu, H. (2006). The cognitive accessibility of synaesthetic metaphors. In R. Sun, & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2365-2370). London: Lawrence Erlbaum.

Whitney, A. H. (1952). Synaesthesia in twentieth-century Hungarian poetry. *The Slavonic and East European Review*, *30*, 444-464.

Whitsitt, S. (2005). A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics*, *10*, 283-305.

Wickham, H. (2017a). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.2.0. https://CRAN.R-project.org/package=strin

Wickham, H. (2017b). tidyverse: Easily install and load 'tidyverse' packages. R package version 1.1.1. https://CRAN.R-project.org/package=tidyverse

Wilce, J. M. (2009). *Language and emotion*. Cambridge, UK: Cambridge University Press.

Willander, J., & Larsson, M. (2006). Smell your way back to childhood: Autobiographical odor memory. *Psychonomic Bulletin & Review*, *13*, 240-244.

Willems, R. M., Labruna, L., D'Esposito, M., Ivry, R., & Casasanto, D. (2011). A functional role for the motor system in language understanding: evidence from theta-burst transcranial magnetic stimulation. *Psychological Science*, *22*, 849-854.

Willems, R. M., & Francken, J. C. (2012). Embodied cognition: taking the next step. *Frontiers in Psychology*, *3*, 582.

Williams, J. (1976). Synaesthetic adjectives: A possible law of semantic change. *Language*, *52*, 461-478.

Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, *322*, 606-607.

Wilson, E. O. (1998). *Consilience: The unity of knowledge*. New York: Alfred A. Knopf.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625-636.

Wilson, N. L., & Gibbs, R. W. (2007). Real and imagined body movement primes metaphor comprehension. *Cognitive Science*, *31*, 721-731.

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, *4*, 58.

Winter, B. (2011). Pseudoreplication in phonetic research. *Proceedings of the International Congress of Phonetic Science* (pp. 2137-2140). Hong Kong, August 2011.

Winter, B. (2014). Horror movies and the cognitive ecology of primary metaphors. *Metaphor and Symbol*, *29*, 151-170.

Winter, B. (2016). Taste and smell words form an affectively loaded part of the English lexicon. *Language, Cognition and Neuroscience*, *31*, 975-988.

Winter, B. (in press). Synaesthetic metaphors are neither synaesthetic nor metaphorical. In L. Speed, L. San Roque & A. Majid (Eds.), *Perception metaphor*. Amsterdam: John Benjamins.

Winter, B., & Bergen, B. (2012). Language comprehenders represent object distance both visually and auditorily. *Language and Cognition*, *4*, 1-16.

Winter, B., & Matlock, T. (2013). Making judgments based on similarity and proximity. *Metaphor and Symbol*, *28*, 219-232.

Winter, B. & Matlock, T. (2017). Primary metaphors are both cultural and embodied. In B. Hampe (Ed.), *Metaphor: Embodied cognition and discourse* (pp. 99-115). Cambridge, UK: Cambridge University Press.

Winter, B., & Matlock, T., Shaki, S., & Fischer, M. (2015a). Mental number space in three dimensions. *Neuroscience and Biobehavioral Reviews*, *57*, 209-219.

Winter, B., Marghetis, T., & Matlock, T. (2015b). Of magnitudes and metaphors: Explaining cognitive interactions between space, time, and number. *Cortex*, *64*, 209-224.

Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, *179*, 213-220.

Winter, B., Perlman, M., & Matlock, T. (2014). Using space to talk and gesture about numbers: Evidence from the TV News Archive. *Gesture*, *13*, 377-408.

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies*, *18*, 433-454.

Wolff, P., & Gentner, D. (2011). Structure-mapping in metaphor comprehension. *Cognitive Science*, *35*, 1456-1488.

Wnuk, E., & Majid, A. (2014). Revisiting the limits of language: The odor lexicon of Maniq. *Cognition*, *131*, 125-138.

Wyler, S. (2007). Color terms between elegance and beauty: The verbalization of color with textiles and cosmetics. In M. Plümacher, & P. Holz (Eds.), *Speaking of Colors and Odors* (pp. 113-128). Amsterdam: John Benjamins.

Yaxley, R. H., & Zwaan, R. A. (2007). Simulating visibility during language comprehension. *Cognition*, *105*, 229-236.

Yeshurun, Y., & Sobel, N. (2010). An odor is not worth a thousand words: from multidimensional odors to unidimensional odor objects. *Annual Review of Psychology*, *61*, 219-241.

Yu, N. (2003). Synesthetic metaphor: A cognitive perspective. *Journal of Literary Semantics*, *32*, 19-34.

Zald, D. H., Lee, J. T., Fluegel, K. W., & Pardo, J. V. (1998). Aversive gustatory stimulation activates limbic circuits in humans. *Brain*, *121*, 1143-1154.

Zald, D. H., & Pardo, J. V. (1997). Emotion, olfaction, and the human amygdala: amygdala activation during aversive olfactory stimulation. *Proceedings of the National Academy of Sciences*, *94*, 4119-4124.

Zangaladze, A., Epstein, C. M., Grafton, S. T., & Sathian, K. (1999). Involvement of visual cortex in tactile discrimination of orientation. *Nature*, *401*, 587-590.

Zarzo, M. (2008). Psychologic dimensions in the perception of everyday odors: pleasantness and edibility. *Journal of Sensory Studies*, *23*, 354-376.

Zatorre, R. J., Jones-Gotman, M., Evans, A. C., & Meyer, E. (1992). Functional localization and lateralization of human olfactory cortex. *Nature*, *360*, 339-340.

Zhang, Q. (1998). Fuzziness - vagueness - generality - ambiguity. *Journal of Pragmatics*, *29*, 13-31.

Zhong, C. B., & Leonardelli, G. J. (2008). Cold and lonely does social exclusion literally feel cold?. *Psychological Science*, *19*, 838-842.

Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of General Psychology*, *33*, 251-256.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. New York: Addison-Wesley.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.

Zwaan, R. A. (2009). Mental simulation in language comprehension and social cognition. *European Journal of Social Psychology*, *39*, 1142-1150.

Zwaan, R. A., & Pecher, D. (2012). Revisiting mental simulation in language comprehension: Six replication attempts. *PLOS ONE*, *7*, e51382.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, *13*, 168-171.

Zwicky, A. M., & Sadock, J. M. (1975). Ambiguity tests and how to fail them. In J. P. Kimball (Ed.), *Syntax and semantics* (pp. 1-36). New York: Academic Press.